Contributions to the analysis of finite element methods for models in electromagnetism, neutronics and fluid dynamics.

Erell Jamelot

INSTITUT POLYTECHNIQUE DE PARIS

Habilitation à Diriger des Recherches

Contributions to the analysis of finite element methods for models in electromagnetism, neutronics and fluid dynamics

Mémoire d'Habilitation à Diriger des Recherches de l'IP Paris
préparé au CEA/DES/ISAS/DM2S/STMF

Habilitation présentée et soutenue à Saclay, le 26 septembre 2025, par

Erell Jamelot

Composition du jury :

Fleurianne Bertrand, *[examinatrice]*	Professeure, Technische Universität Chemnitz, Allemagne
Xavier Claeys *[rapporteur]*	Professeur, ENSTA Paris
Sonia Fliss *[présidente du jury]*	Professeure, ENSTA Paris
Olga Mula-Hernandez *[rapporteure]*	Professeure, University of Vienna, Autriche
Serge Nicaise *[examinateur]*	Professeur, Université Polytechnique Hauts-de-France
Stéphanie Salmon *[rapporteure]*	Professeure, Université de Reims-Champagne Ardenne
Claire Scheid *[examinatrice]*	Maîtresse de conférence, Université Côte d'Azur

Résumé

La méthode des éléments finis est fréquemment utilisée pour résoudre des équations aux dérivées partielles issues de problèmes industriels. Après avoir soutenu une thèse qui portait sur la discrétisation des équations de Maxwell, j'ai contribué à l'analyse numérique, à la simulation et au développement de codes du CEA pour résoudre des problèmes issus de la neutronique et de la mécanique des fluides. Ces différents modèles ont des caractéristiques en commun : il s'agit de problèmes elliptiques dont la solution peut être de régularité faible. De plus, ils sont posés sous une forme mixte, soit à cause d'une contrainte (équations de Maxwell et problème de Stokes), ce qui conduit à un problème de point-selle, soit pour améliorer la discrétisation du gradient (équation de transport simplifiée des neutrons). Enfin, pour les équations de Maxwell et pour l'équation de transport simplifiée des neutrons, nous nous intéressons également à la résolution des problèmes propres. Dans ce manuscrit, pour ces trois modèles, nous procédons de manière similaire : étude du modèle physique continu, puis, lorsque cela est possible, analyse numérique des méthodes d'éléments finis. Enfin, on propose des algorithmes numériques. Le cas échéant, nous utiliserons un outil mathématique récent : la théorie de la T-coercivité.

Abstract

The finite element method is frequently used to solve partial differential equations arising from industrial problems. After defending a PhD thesis on the discretisation of Maxwell's equations, I contributed to numerical analysis, simulation and the development of CEA codes to solve problems arising from neutronics and fluid mechanics. These different models have characteristics in common: they are elliptic problems whose solution may be of low regularity. In addition, they are posed in a mixed form, either because of a constraint (Maxwell's equations and Stokes' problem), which leads to a saddle point problem, or to improve the discretisation of the gradient (simplified neutron transport equation). Finally, for Maxwell's equations and the simplified neutron transport equation, we are also interested in solving eigenvalue problems. In this manuscript, we proceed in a similar manner for these three models: study of the continuous physical model, then, where possible, numerical analysis of finite element methods. Finally, we propose numerical algorithms. Where appropriate, we will use a recent mathematical tool: the T-coercivity theory.

Remerciements

Ce travail n'aurait pas été possible sans l'engagement du CEA en faveur de la Recherche. Je suis extrêmement reconnaissante envers le CEA de m'avoir soutenue dans mon choix de poursuivre la filière d'expert scientifique et de préparer une HDR. Je remercie donc Philippe Prené, chef de l'ISAS et Stéphane Gossé, adjoint scientifique de l'ISAS; Lydie Grospellier, cheffe du DM2S et Daniel Caruge, ajoint scientifique du DM2S; Nicolas Dorville, chef du STMF et Philippe Fillion, adjoint scientifique du STMF. Je remercie également Cheikh Diop, adjoint scientifique du SERMA. Merci à Julie Darona, cheffe du LMSF puis du LDEL pour son soutien et ses conseils. Merci à Maria Giovanna Rodio, cheffe du LCAN pour son invitation au projet Innoflex. Enfin, je remercie, Katy Colafrancesco, secrétaire du LDEL pour l'aide à l'organisation de ma soutenance.

Le CEA fait partie de l'Université Paris-Saclay, mais comme l'essentiel de mon activité de recherche est davantage associée à l'Institut Polytechnique de Paris qu'à un autre établissement de recherche, j'ai demandé à m'inscrire à l'IP de Paris pour soutenir une HDR. Je remercie donc Raphaël Krikorian, référent HDR IP de Paris pour les mathématiques, pour avoir soutenu ma démarche, pour le suivi scientifique de mon dossier, pour ses conseils et ses encouragements. Merci à Emmanuel Fullenwarth pour le suivi administratif.

Du côté de l'École Doctorale de Mathématiques Hadamard, je remercie Agnès Desolneux, la directrice de l'EDMH et Éliane Bécache, son adjointe pour nous avoir permis d'inscrire Paulin Dempowo-Fouelefack en thèse à l'IP de Paris, bien que le CEA soit affilié à l'Université Paris-Saclay.

Je tiens à exprimer ma profonde gratitude envers Xavier Claeys, Stéphanie Salmon et Olga Mula pour avoir accepté de consacrer du temps à l'évaluation de ce manuscrit. Leurs remarques constructives, leur expertise et leur confiance m'ont été très précieuses. Je remercie très chaleureusement Fleurianne Bertrand, Sonia Fliss, Serge Nicaise et Claire Scheid d'avoir accepté de participer à mon jury, et d'avoir prêter attention à mon travail. Merci à Sonia Fliss d'avoir pris la responsabilité d'être présidente du jury. Merci à Fleurianne, Serge, Stéphanie et Claire d'être venu "en présentiel". Chers membres du jury, c'était un véritable honneur que vous ayez tous accepté de m'évaluer. J'ai passé vraiment un excellent moment, je m'en souviendrai longtemps ! Merci à Fleurianne pour les discussions après la soutenance. Merci au passage à Alain Cartalade le covoiturage !

Ce travail n'aurait pas pu se faire non plus sans collaborations. Je tiens à remercier en particulier Patrick Ciarlet de l'ENSTA pour sa confiance de longue date, ses conseils, suggestions et enseignements; et pour avoir diriger la thèse de Léandre Giret. Un grand merci à Olivier Lafitte de l'Institut Galilée, pour notre collaboration sur le couplage neutronique-thermohydraulique, pour son enthousiasme et pour me faire sortir de ma "zone de confort". Je remercie Éric Chénier, et Christophe Le Potier pour avoir co-dirigé la thèse d'Andrew Peitavy. Merci également à Pascal Omnes, pour la direction de thèse de Mayssa Mroueh et la co-direction de thèse de Paulin Dempowo. Chers anciens et futurs doctorants : Léandre Giret et Andrew Peitavy; Mayssa Mroueh et Paulin Dempowo, merci pour la confiance que vous m'avez accordé (ou que vous m'accordez) pendant votre thèse.

C'est une activité très enrichissante que d'enseigner, et j'ai eu la chance de pouvoir le faire : grâce à Patrick Ciarlet, qui m'a permis de participer pendant 5 ans au cours de calcul scientifique du M2 AMS ; et à Olivier Lafitte et Pascal Omnes qui m'ont proposé de donner un cours "Projet Métier" en deuxième année de l'école d'ingénieurs Sup Galilée pendant 3 ans. Merci de m'avoir offert ces opportunités.

Je remercie mes co-auteurs : Patrick Ciarlet, Éric Chénier, Olivier Lafitte, Christophe Le Potier, Pascal Omnes, Pierre-Emmanuel Angeli et François Madiot. Many thanks also to Stefan Sauter, for all these discussions about the nonconforming Crouzeix-Raviart finite elements! Nos échanges stimulants, votre soutien et votre amitié m'aident à garder le cap scientifique, à m'instruire et à évoluer. Merci également à Nathalie Nouaime et Maria Adela Puscas.

Merci beaucoup aux collègues, permanents ou non, pour leurs attentions, et les échanges scientifiques ou autre. Merci enfin au groupe de promeneurs pour m'inciter à faire de temps en temps les dix mille pas journaliers recommandés !

Contents

Part I

Résumé

Chapter 1

Publications and communications

Some of my work has been more theoretical in nature, related to numerical analysis, and developed through articles and conference presentations. Another part concerns the application of research to industrial codes, which is developed through conference proceedings and technical reports. As I work as an engineer-researcher, this part contains more publications. I also took part in an M2 course in scientific and parallel computing, which led me to participate in the writing of a handout.

1.1 Publications

Published articles

Below is a list of my published or accepted articles, from the most recent to the oldest.

[A13] *Description and convergence order analysis of the Finite Element-Volume spatial discretization method*, M. A. Puscas, P.-E. Angeli, N. Nouaime, E. Jamelot, International Journal for Numerical Methods in Fluids, 2025.

[A5] *Explicit T-coercivity for the Stokes problem: a coercive finite element discretization*, P. Ciarlet Jr, E. Jamelot, Computers & Mathematics with Applications, **188**, pp. 137–159, 2025.

[A9] *Stability estimates for solving Stokes problem with nonconforming finite elements*, E. Jamelot, Comptes Rendus. Mathématiques, **363**, pp. 115-137, 2025.

[A4] *Variational methods for solving numerically magnetostatic systems*, P. Ciarlet Jr, E. Jamelot, Advances in Computational Mathematics **50**, 5, 2024.

[A12] *Numerical analysis of the neutron multigroup SPN equations*, E. Jamelot, F. Madiot, Comptes Rendus Mathématique, **359**, 5, pp. 533–545, 2021.

[A2] *Numerical analysis of the mixed finite element method for the neutron diffusion problem with highly heterogeneous coefficients*, P. Ciarlet Jr, L. Giret, E. Jamelot, F. Kpadonou, ESAIM: Mathematical Modelling and Numerical Analysis, **52**, 5, pp. 2003–2035, 2018.

[A6] *Domain decomposition methods for the diffusion equation with low-regularity solution*, P. Ciarlet Jr, E. Jamelot, F. Kpadonou, Computers and Mathematics with Applications, **74**, pp. 2369–2384, 2017.

[A7] *A simple monodimensional model coupling an enthalpy production equation and a neutronic diffusion equation*, S. Dellacherie, E. Jamelot, O. Lafitte, Applied Mathematics Letters, **62**, pp. 35–41, 2016.

[A11] *Fast non-overlapping Schwarz domain decomposition methods for solving the neutron diffusion equation*, E. Jamelot, P. Ciarlet Jr, Journal of Computational Physics, **241**, pp. 445–463, 2013.

[A10] *Domain Decomposition for the SPN Solver MINOS*, E. Jamelot, A.-M. Baudron, J.-J. Lautard, Transport Theory and Statistical Physics, **41** pp. 495–512, 2012.

[A1] *Solving electromagnetic eigenvalue problems in polyhedral domains with nodal finite elements*, A. Buffa, P. Ciarlet Jr, E. Jamelot, Numerische Mathematik, **113**, 4, pp. 497–518, 2009.

[A3] *Continuous Galerkin methods for solving the time-dependent Maxwell equations in 3D geometries*, P. Ciarlet Jr, E. Jamelot, Journal of Computational Physics **226**, 1, pp. 1122–1135, 2007.

[A8] *Éléments finis nodaux pour les équations de Maxwell*, E. Jamelot, C. R. Acad. Sci. Paris, Ser. I, **339**, 11, pp. 809–814, Corrigendum C. R. Acad. Sci. Paris, Ser. I, vol. **340**, 5, pp. 409–410, 2005, déc 2004.

Peer-reviewed conference proceedings

Below is a list of my conference proceedings, from the most recent to the oldest.

[P7] *Stability of the* $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ *element*, E. Jamelot, P. Ciarlet Jr., and S. Sauter, ENUMATH, Lisbonne 2023, Sept. 4-8, Lisbonne (Portugal).

[P9] *Discontinuous Galerkin schemes for the Stokes problem*, M. Mroueh, E. Jamelot, P. Omnes, SNA + MC 2024, Oct. 20-24, Paris.

[P4] *Improved Crouzeix-Raviart scheme for the Stokes and Navier-Stokes problem*, É. Chénier, E. Jamelot, C. Le Potier, A. Peitavy, New trends in complex flows, 2022, Sept. 19-21, Institut Henri Poincaré, Paris.

[P3] *Improved Crouzeix-Raviart scheme for the Stokes problem*, É. Chénier, E. Jamelot, C. Le Potier, A. Peitavy, Finite Volumes for Complex Applications **10**, 2023, Oct. 30-Nov. 3, Strasbourg.

[P5] *Critical Computation with Finite Element Method on Non-Conforming Meshes*, L. Giret, P. Ciarlet Jr, E. Jamelot, M&C 2017, April 16-20, Jeju (Korea).

[P2] *Numerical results for the coupling of a simple neutronics diffusion model and a simple hydrodynamics low mach number model without coupling codes*, S. Dellacherie, E. Jamelot, O. Lafitte, R. Mouhamad, SYNASC 2016, September 24-27, Timişoara (Romania).

[P10] *APOLLO3®: CEA/DEN deterministic multi-purpose code for reactor physics analysis*, D. Schneider et al, PHYSOR 2016, May 1-5, Sun Valley (Idaho, USA).

[P6] *Domain decomposition for the neutron SPN equations*, E. Jamelot, P. Ciarlet Jr., A.-M. Baudron, J.-J. Lautard, DDM'21, 2012, June 25-29, Rennes.

[P8] *High performance 3D neutron transport on petascale and hybrid architectures within APOLLO3® code*, E. Jamelot, J. Dubois, J.-J. Lautard, C. Calvin, A.-M. Baudron Proceedings of M&C 2011, May 8-12, Rio de Janeiro (Brazil).

[P1] *Continuous Galerkin methods for solving Maxwell equations in 3D geometries*, P. Ciarlet Jr., E. Jamelot, ENUMATH 2005, July 18-22, Santiago de Compostela (Spain).

Articles submitted or soon to be finalised

[W1] *Convergence of the* $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ *discretization in the TrioCFD code*, P.-E. Angeli, H. Bertrand, E. Jamelot, submitted to Computers & Mathematics With Application.

[W2] *A Priori Error Estimates for Discontinous Galerkin Finite Element Discretization of Stokes Problems with Low Regularity*, E. Jamelot, M. Mroueh, P. Omnes.

Lecture notes, level M2 and PhD thesis (in French)

[B1] *Outils mathématiques et algorithmiques pour le calcul scientifique* (288 pages), P. Ciarlet Jr, E. Jamelot, courses included in the program Modelisation and Simulation of M2 Analysis, Modeling, Simulation at Université Paris-Saclay and in the program ModSim, last year of ENSTA Paris, 2019.

[B2] *Continuous Galerkin methods for solving Maxwell's equations*, Erell Jamelot, PhD thesis, École Polytechnique, Palaiseau, France, director: Patrick Ciarlet (2005).

GitHub digital models to solve Stokes problem in $2D$

[G2] *Explicit* T*-coercivity to solve Stokes problem*, E. Jamelot, 2024.

[G1] *Non-conforming finite elements to solve Stokes problem*, E. Jamelot, 2022.

CEA monograph and technical reports

2022 *Les éléments finis Galerkin discontinus pour le problème de Stokes*, E. Jamelot, DES/ISAS/DM2S/STMF/LMSF/NT/2022-71066/A, 2022.

2020 *Neutronique et calcul intensif*, C. Diop et al, Les sciences du numérique et le calcul haute performance, pp. 187-210, 2020.

2018 *Méthodes d'éléments finis de Galerkin discontinus pour résoudre les équations de Navier-Stokes*, E. Jamelot, DEN/DANS/DM2S/STMF/LMSF/NT/2018-64097/A, 2018.

2018 *A proposal to improve the stability of the* $\mathbf{P}^{nc}_1 \times P_0$ *scheme in the TrioCFD code*, E. Jamelot, DEN/DANS/DM2S/STMF/LMSF/NT/2018-62465/A, 2018.

2017 *Différents modèles analytiques ou simplifiés pour le couplage Thermohydraulique-Neutronique*, E. Jamelot, O. Lafitte, J.-C. Le Pallec, DEN/DANS/DM2S/SERMA/LLPR/NT/17-6134/A, 2017.

2015 *Décomposition de domaine pour le solveur MINOS du code APOLLO3®*, E. Jamelot, A.-M. Baudron, J.-J. Lautard, DEN/DANS/DM2S/SERMA/LLPR/NT/15-5934/A, 2015.

2011 *Résultats de la parallélisation du solveur MINOS dans APOLLO3® appliquée à un calcul de PN*, E. Jamelot, A.-M. Baudron, J.-J. Lautard, DEN/DANS/DM2S/SERMA/LLPR/RT/11-5154/A, 2011.

2009 *Domain decomposition method for core calculations using MINOS solver in the APOLLO3® code*, E. Jamelot, J.-J. Lautard, DEN/DANS/DM2S/SERMA/LLPR/RT-4755/A, 2009.

Popularizing science (in French)

2021 *Les équations de Navier-Stokes*, presentation for the *Fête de la Science*, organized by the Académie de Paris for high-school science students.
E. Jamelot, P. Omnes, La Sorbonne, Paris, 4 octobre 2021.

2021 *Les équations de Navier-Stokes*, video #7 of the Key Equations of Physics cycle *Les équations Clefs de la Physique*). Scientific advisors : E. Jamelot, P. Omnes.

2020 *Simuler les écoulements*, article from the CEA magazine *Clefs #70*,
N. Dorville, E. Jamelot, P. Omnes.

1.2 Oral communications

Invited talks

2025 *Explicit T-coercivity for the Stokes problem: a coercive finite element discretization*, E. Jamelot, P. Ciarlet Jr, Olivier Lafitte's 60th birthday, École des Mines, 2025, June 30-July 2, Paris.

2024 *The T-coercivity Approach for Solving Stokes Problem*, E. Jamelot, P. Ciarlet Jr,
ECCOMAS (9th European Congress on Computational Methods in Applied Sciences and Engineering), 2024, June 3-7, Lisbon (Portugal).

2019 *TrioCFD: code & numerical schemes*, E. Jamelot et al,
Conference POEMS, CIRM, 29 avril-3 mai 2019.

2017 *Nuclear core reactor simulations with low-regularity solution*, P. Ciarlet Jr, L. Giret, E. Jamelot, F. D. Kpadonou, Rencontres Inria-LJLL en calcul scientifique, 15 mai 2017.

2016 *Domain Decomposition Methods with low-regularity solution for nuclear core reactor simulations*, P. Ciarlet Jr, L. Giret, E. Jamelot, F. D. Kpadonou,
Eigth Singular Days, Nancy (France), 27-30 Juin 2016.

2012 *Calculs haute performance en diffusion et transport neutronique*, A.-M. Baudron, C. Calvin, J. Dubois, E. Jamelot, J.-J. Lautard, O. Mula-Hernandez, Forum Horizon Maths, Fondation Science Mathématiques de Paris, La Défense (France), 20 décembre 2012.

2004 *Résolution des équations de Maxwell dans des domaines singuliers*,
P. Ciarlet Jr, E. Jamelot, Fourth Singular Days, Pont-à-Mousson (France), 2004.

Participation to workshop specialized in nuclear science

2021 *Étude de la T-coercitivité pour le problème de Stokes*,
P. Ciarlet Jr, E. Jamelot, A. Lefort, Journées du GdR MaNu, Le Croisic, 25-27 octobre 2021.

2018 *TrioCFD: code & numerical schemes*, E. Jamelot,
Workshop DAM/DEN HPC, Aussois, 2-6 juillet 2018.

2017 *Analysis of the mathematical tools to perform nuclear core reactor simulations*,
P. Ciarlet Jr, L. Giret, E. Jamelot, F. D. Kpadonou,
Workshop "Mathématiques pour la neutronique", Groupe de Recherche MaNu, LJLL, Université Paris 6, 30 novembre 2017.

2017 *Un modèle simplifié de couplage thermohydraulique-neutronique*,
S. Dellacherie, E. Jamelot, O. Lafitte, R. Mouhamad,
Journée scientifique autour de la thermohydraulique, programme NEEDS,
MAP5, Université Paris 7, 5 juillet 2017.

2017 *Couplage entre la thermohydraulique et la neutronique*,
S. Dellacherie, E. Jamelot, O. Lafitte, R. Mouhamad, Atelier FrenchTeam-MSFR,
programme NEEDS, Lyon, 2 février 2017.

2016 *Domain Decomposition Methods with low-regularity solution for nuclear core reactor simulations*,
P. Ciarlet Jr., L. Giret, E. Jamelot, F. D. Kpadonou,
Groupe de Recherche MaNu, Saint-Valéry-sur-Somme, 17-19 octobre 2016.

2015 *Décomposition de domaine non conforme pour l'équation mixte de la diffusion avec une solution peu régulière*, P. Ciarlet Jr., E. Jamelot, L. Giret, LRC MANON, Paris, 16 juin 2015.

2015 *Calcul neutronique déterministe des réacteurs avec prise en compte des contre réactions simplifiées*,
E. Jamelot, J.-J. Lautard, J.-C. Le Pallec, LRC MANON, Paris, 2 juin 2015.

2014 *Décomposition de domaine pour la neutronique*, A.-M. Baudron, E. Jamelot, J.-J. Lautard,
LRC MANON, Paris, 12 juin 2014.

2011 *Parallélisation du solveur de transport simplifié MINOS dans APOLLO3®*,
A.-M. Baudron, E. Jamelot, J.-J. Lautard, LRC MANON, Paris, 17 juin 2011.

Organization of scientific events

2025 *Workshop on numerical analysis and scientific computing for electromagnetics.* Patrick Ciarlet's 60th birthday, Institut Henri Poincaré, Paris, 17 juin 2025.

2024 *New Directions for solving Stokes and Navier-Stokes Problems.* Minisymposium MS061, ECCOMAS (9th European Congress on Computational Methods in Applied Sciences and Engineering) 2024, June 3-7, Lisbon (Portugal).

2022 Participation in the organization committee of the CEA-EDF-INRIA summer school on numerical analysis.

2017-now Organization of seminaries, PhD proposals.

2014-2017 SERMA correspondent for the *Laboratoire de Recherche Conventionné* (LRC) MANON, organization of seminaries, PhD and post-doc proposals.

Contributed talks without proceedings

2023 *Stability estimates for Fortin-Soulie mixed finite elements*, E. Jamelot,
11ème Biennale de la SMAI, Le Gosier (France), 22-26 mai 2023.

2022 *Étude de la T-coercitivité pour le problème de Stokes*, E. Jamelot, A. Lefort,
CANUM, Evian (France), 13-17 juin 2022.

2020 *Domain Decomposition Method for nuclear core reactor simulations with low-regularity solution*, P. Ciarlet Jr., L. Giret, M. H. Do, E. Jamelot, F. D. Kpadonou, F. Madiot,
26th International Domain Decomposition Conference, 2020, December 7-12 (Hong Kong).

2019 *Eigenproblem with low-regularity solution for nuclear reactor core modeling*,
P. Ciarlet, Jr., L. Giret, E. Jamelot, F. D. Kpadonou,
MAFELAP'19, 2019, June 18-21, Brunel (UK).

2016 *Domain Decomposition Methods with low-regularity solution for nuclear core reactor simulations*, P. Ciarlet Jr., L. Giret, E. Jamelot, F. D. Kpadonou, MAFELAP'16, 2016, June 14-17 Brunel (UK).

2010 *GPGPU Programming to Solve the Boltzman Neutron Transport Equation*, C. Calvin, J. Dubois, E. Jamelot, S. Petiton, 14^{th} SIAM Conference on Parallel Processing for Scientific Computing, Seattle (USA), Feb. 2010.

2006 *Continuous Galerkin methods for solving electromagnetic eigenvalue problems*, A. Buffa, P. Ciarlet Jr., E. Jamelot, MAFELAP'06, 2006, June 13-16, Brunel (UK).

2004 *Résolution des équations de Maxwell dans des domaines singuliers*, P. Ciarlet Jr., E. Jamelot, CANUM, Obernai (France), juin 2004.

2003 *Augmented variational formulations for solving Maxwell's equations: numerical results*, P. Ciarlet Jr., E. Jamelot, MAFELAP'03, 2003, June 21-24, Brunel (UK).

Chapter 2

Teaching and student supervision

I started teaching at ENSTA and Université Paris-Dauphine during my PhD. For 5 years, I was involved in teaching the M2 AMS course (ENSTA and Université Paris-Saclay), for the scientific computing course. For the last two years, I taught at Université Sorbonne Paris Nord, for M1 students. I also supervise or co-supervise M1, M2 and PhD students.

2.1 Teaching

2025–... TD et TP, *Méthode des éléments finis*, ENSTA, 15 hours (M1).

2023–2025 *Introduction à la résolution des équations de Navier-Stokes*,
École d'Ingénieurs Sup'Galilée (Université Sorbonne Paris Nord), 15 hours.
Cours "Projet métier" de la Spécialité Mathématiques Appliquées au Calcul Scientifique (M1).

2014–2019 *Calcul scientifique parallèle*, ENSTA, 15 hours.
Cours du Master Analyse, Modélisation et Simulation (M2).

2012–2014 TD et TP, *Méthode des éléments finis*, ENSTA, 15 hours (M1).

2008 Qualifiée aux fonctions de maître de conférences en section 26 du CNU.

2007–2008 TD et TP, *Méthode des éléments finis*, ENSTA, 15 hours (M1).

2003–2005 Cours-TD, *Analyse fonctionnelle*, Université Paris-Dauphine (L1 GEA).

2003–2005 TP, *Programmation en Matlab*, ENSTA (L3).

2.2 Students supervision

PhD students

2025–2028 **Paulin Dempowo-Fouelefack**, Université Paris-Saclay.
Méthodes numériques pour la simulation d'écoulements diphasiques avec prise en compte des singularités. Director: Erell Jamelot (34%). Co-instructors: P. Omnes (33%) and M. Grosso (33%).

2022–2025 **Mayssa Mroueh**, Université Sorbonne Paris-Nord.
Étude de la méthode de Galerkine discontinue pour les équations de Navier-Stokes.
Director: Pascal Omnes (50%). Proceedings [P9], manuscript in preparation [W2].

2020–2024 **Andrew Peitavy**, Université Gustave Eiffel.
Schémas de discrétisation en pression et éléments finis de Crouzeix-Raviart pour les écoulements de fluides incompressibles. Directors: Éric Chénier (25%) and Christophe Le Potier (25%). Proceedings [P3, P4], PhD thesis [210].

2014–2018 **Léandre Giret**, Université Paris-Saclay.
Numerical Analysis of a Non-Conforming Domain Decomposition for the Multigroup SPN Equations. Director: Patrick Ciarlet (50%). Article [A2], proceedings [P5], PhD thesis [163].

Work-study student

10/22–09/24 **Paulin Dempowo-Fouelefack**, ENSTA.
Méthodes numériques et développement de code CFD diphasique.
Co-instructor with Raksmy Nop and Pascal Omnes.

Gap year students

09/22–02/23 **Sofiane Ezzehi**, ENPC.
Approche semi-analytique pour un couplage simplifié thermohydraulique-neutronique.
Co-instructor with Olivier Lafitte and François Madiot.

02/21–07/21 **Albéric Lefort**, ENSTA. *Étude de la T-coercivité pour le problème de Stokes.*

M2 students

04/25–09/25 **Raphaël Lecoq**, Sciences Sorbonne Université.
Estimateurs d'erreurs a posteriori pour les éléments finis de Galerkin discontinus.
Co-instructor with Andrew Peitavy.

04/24-09/24 **Honorine Bertrand**, Université de Reims Champagne-Ardennes.
Analyse numérique du schéma $\mathbf{P}^1_{nc} - (P^0 + P^1 + P^a)$ *du code TrioCFD.*
Co-instructor with Pierre-Emmanuel Angeli. Submitted article [W1].

04/23–09/23 **Ali Ali Ahmad**, Université de Reims Champagne-Ardennes.
Étude des schémas de convection du code TrioCFD. Co-instructor with Pierre-Emmanuel Angeli.

04/22–09/22 **Mayssa Mroueh**, Université Versailles Saint-Quentin-en-Yvelines.
Étude de la T-coercitivité pour le problème de Stokes discrétisé avec les éléments finis de Galerkine discontinus. Co-instructor with Patrick Ciarlet.

04/20–11/20 **Andrew Peitavy**, Centrale Marseille. *Méthode hybride d'ordre élevée pour l'équation de Stokes.*

03/19–09/19 **Loria Dray**, Université Pierre et Marie Curie.
Schémas numérique à base de splines pour l'équation de Stokes.

03/18–09/18 **Mahran Rihani**, ENSTA Paris. *Etude d'un élément fini mixte d'ordre 2 et correction de la faible conservation de la masse pour l'équation de Stokes.*

04/13–08/13 **Félix D. Kpadonou**, Université Versailles Saint-Quentin-en-Yvelines.
Études de méthodes de décomposition de domaine sur maillages non conformes.
Co-instructor with Patrick Ciarlet. Article [A6].

03/05–07/05 **Mathieu Buard**, Université du Maine.
Résolution des équations de Maxwell en $2D$. Co-instructor with Patrick Ciarlet.

M1 students

05/14–08/14 **Hervé Avezedo-Chavez**, Université Sorbonne Paris Nord.
Implémentation du GMRES pour la résolution de la méthode de décomposition de domaine du solveur MINOS du code APOLLO3®.

03/09–07/09 **France Boillod-Cerneux**, Polytech Lille.
Utilisation de la bibliothèque PETSc dans le solveur MINOS du code APOLLO3®.
Co-instructor with Christophe Calvin.

2.3 Examiner

08/31/17 Examiner on the jury for Félix Kpadonou's thesis [181].

2014–2019 Examiner on the jury for the M2 AMS (Université Paris-Saclay).

Chapter 3

Abstract of the contributions

As this manuscript concerns my research work, let's first check what a famous AI language model says about me: *Erell Jamelot is a research engineer affiliated with the Université Paris-Saclay and the CEA (French Alternative Energies and Atomic Energy Commission). Her work focuses on applied mathematics and numerical analysis, with expertise in the finite element method, computational physics, and mathematical modeling of physical systems. She has contributed to research on problems such as Maxwell equations, and numerical methods for fluid mechanics and neutron transport equation.* That's quite right, as I'll explain in detail (without the AI).

I defended my PhD thesis in November 2005, on *Continuous Galerkin methods for solving Maxwell's equations* (cf. manuscript [B2], article [A8] and proceedings [P1]). During my PhD, I developed a vector finite element code, implementing different vector continuous finite elements: the singular complement method in $2D$ [173, 157], the method with natural boundary conditions [91] and the weighted regularization method in $2D$ and $3D$ [116].

In January 2006, I started working at Dassault Systèmes, where I took part in the development of the well-known CAD software, CATIA. As I wanted to take part in a subject closer to physics, in January 2008, I joined the Phasics start-up, where I developed algorithms for adaptive optics. However, while I held these two positions, I wasn't doing research, and I missed it. That's what led me to apply to the CEA, whose motto is *From Research to Industry*[1]. Since then I've tried to follow this motto, collaborating with academic researchers, including: Patrick Ciarlet, Olivier Lafitte, Eric Chénier and CEA colleagues, including: François Madiot, Christophe Le Potier, Pascal Omnes, Pierre-Emmanuel Angeli, with the aim of improving the performance of CEA's industrial codes. Last, since 2022, I started a collaboration with Stefan Sauter on nonconforming finite elements.

I was recruited in October 2008 at the Reactors studies and Applied Mathematics Department (SERMA[2]). There, I took part in the development of finite element methods to solve the neutron transport equation in the APOLLO3® code[3] (cf. articles [A12, A2, A6, A11, A10] and proceedings [P5, P10, P6, P8]). Since then, I co-supervised the PhD thesis of Léandre Giret [163]. From June 2015, I started collaborating with Olivier Lafitte on a simplified thermal-hydraulics and neutronic coupling (cf. article [A7] and proceedings [P2]), which led me, two years later, to join the Thermohydraulics and Fluid Mechanics department (STMF[4]).

Since April 2017, my research has focused mainly on the spatial discretization of the Navier-Stokes equations and the Stokes problem, in order to enhance capabilities of CEA's open source computational fluid dynamics (CFD) code, called TrioCFD (articles [A5, A9, A13], proceedings [P3, P4, P9, P7]. I co-supervised the PhD thesis of Andrew Peitavy [210]. Currently, I'm co-supervising the PhD thesis of

[1] *De la recherche à l'industrie.*
[2] *Service d'Étude des Réacteurs et de Mathématiques Appliquées.*
[3] APOLLO3 is a registered trademark in France.
[4] *Service de Thermohydraulique et de Mécanique des Fluides.*

Mayssa Mroueh, who studies the discretization of Navier-Stokes equations with discontinuous Galerkin finite elements (cf. proceedings [P9] and article [W2]). I started to co-supervive the PhD thesis of Paulin Dempowo-Fouelefack on the modelisation and simulation of two phases flows.

Thus, I mainly studied **three physical models: Maxwell's equations, the neutron transport equation and the Stokes problem.** These models have some features in common: they are elliptical problems. Moreover, they are posed in a mixed form, either because of a constraint (Maxwell's equations and Stokes problem), which leads to a saddle point problem, or to improve the discretization of the gradient (the neutron simplified transport equation). Finally, for Maxwell's equations and for the neutron simplified transport equation, we are also interested in solving eigenproblems.

In this manuscript, for these three models, we will proceed in a similar way: we will study the continuous, physical model, then, when possible, we will carry out the numerical analysis of the finite element methods. Last, we may propose numerical algorithms. **When relevant, we will use a recent mathematical tool: the T-coercivity theory [95].** On the one hand, it can help proving well-posedness of continuous or discrete problems (basic or discrete T-coercivity). On the other hand, it can be used to demonstrate stabilization or preconditioning (explicit T-coercivity).

The manuscript is organized as follows: in **Chapter 4**, we gather the **mathematical tools** that we need for the mixed formulations, T-coercivity and low-regular solutions. Chapters 5-7 are dedicated to each of the three models. Last, in Chapter 8, we propose some perspectives. Some details are developed in appendix A and B. Below, we give precisions on Chapters 5-7.

My thesis was followed by articles [A1, A3], and I still work occasionally on Maxwell equations to review articles or to answer questions from colleagues. This is the reason why recently, Patrick Ciarlet proposed to collaborate on solving magnetostatic systems (cf. article [A4]). These articles are summarized in **Chapter 5**, where we study three **electromagnetism models**: magnetostatic systems in Sections 5.2 and 5.4; the electric eigenproblem in Section 5.5 and the propagation in a waveguide in §5.1.

In the most recent article [A4, §5], we design a **new variational formulation with the help of T-coercivity**, adapted to magnetostatic systems, as explained in Section 5.4.1. This formulation can be discretized with edge finite elements for which we carry out the numerical analysis.

In order to be able to solve Maxwell's equations in a **non-convex polyhedron** with a **continuous and conforming discretization**, one has to overcome a very difficult mathematical problem. It turns out that the discretized spaces are always included in a closed, strict subspace (sometimes called the subspace of regular fields) of the space of all possible fields. In other words, one cannot hope to approximate the part of the field, if it exists, which belongs to the orthogonal of this subspace. Several methods have been devised to address this problem. In particular, one can **use a weaker norm to measure the divergence of the field**: one replaces the L^2-norm by a weighted L^2-norm. This weighted approach, introduced by Costabel and Dauge [116, 117], leads again to a larger space for the fields, where the density property is recovered.

We describe the **weighted regularization method (WRM)** in Section 5.3. We study the resolution of magnetostatic systems [A4, §6] with the WRM in Section 5.4.2. In Sections 5.5 and 5.6, we study the electric eigenproblem [A1] and of the propagation in a waveguide [A3].

From the end of 2008 to March 2017, I worked at SERMA on the development of methods for solving **the neutron transport equation**, used to simulate nuclear reactor cores. This work has given rise to articles [A10, A11, A6, A2, A12] and proceedings [P8, P6, P5], which are summarized in **Chapter 6**.

The coefficients of the elliptic operator are piecewise regular. Thus, the solution to the neutron transport equation (called the neutron flux) may be of **low-regularity**. The steady state of a nuclear reactor core is modelled by an **eigenproblem**, whose principal mode is the physical solution to the neutron transport equation. The neutron transport equation depends on 7 variables (the time, the velocity of the neutrons and the position), the approximation of which leads to different models, depending on how

the direction of the velocity is discretized. In Section 6.1, we describe our main model: the neutron simplified transport equation (also called neutron SP_N equations), which corresponds to a set of coupled diffusion equations, and the neutron diffusion equation.

Pressurized water reactors have a Cartesian geometry, therefore, it is interesting to discretize the model with Cartesian **Raviart-Thomas-Nédélec (RTN) finite elements**. In Sections 6.2 and 6.3, we carry out the numerical analysis of the neutron diffusion source problem, and eigenproblem discretized with RTN finite elements.

To take advantage of new computer architectures and speed up calculations, we need to work on **parallel methods**. Thus, we are interested in **domain decomposition methods** (DDM) to solve the simplified neutron transport equations, discretized with **RTN finite elements** [213, 204]. In Section 6.4, we introduce DDM and related notations. In Section 6.5, we give details on **the Schwarz DDM with optimised Robin boundary conditions**, cf. articles [A11, A10], proceedings [P5, P6] and lecture notes [B1, chap. 3]. This method is efficient, but it cannot handle nonconforming meshes. For this reason, we also studied **nonconforming DDM** with Lagrange multipliers (or joints), as explained in Section 6.6. In order to carry out the numerical analysis for **low-regular solutions**, we make use of **vector-valued functions with L^2-jumps of normal traces on the interface between subdomains** that we called the **DD+L^2-jumps method** cf. article [A6]. This research led me to supervise Léandre Giret's PhD thesis [163], cf. article [A2] and proceedings [P5].

Discontinuous Galerkin finite element method (DGFEM) was first designed to solve the neutron transport equation [215, 192]. Then, this method was adapted to elliptic problems. The diffusion equation can also be used as preconditioner to solve the transport equation. In Section 6.7, we study the discretization of the simplified neutron transport equations, with the so-called **symmetric interior penalty method** [A12].

Last, in Section 6.8, we study **a simple neutronics and thermal-hydraulics coupling** cf. article [A7], proceedings [P2]. We obtain an analytic solution of a monodimensional stationary system coupling simplified thermal-hydraulics and neutronic models based on the diffusion approximation with one energy group. In principle, the procedure we propose could be used to validate other coupling strategies used for more realistic models.

In **Chapter 7**, we discuss recent work on fluid mechanics, done at the STMF since April 2017. We present the Navier-Stokes equations and the Stokes problem in Section 7.1. From a numerical analysis point of view, the study of the Stokes problem represents an essential step in constructing a good spatial discretization for the Navier-Stokes equations, and we concentrate on this problem. In Section 7.2, we study the **well-posedness of Stokes problem with the help of T-coercivity**. Next, in Section 7.3, we introduce useful discrete notations. In TrioCFD code, the spatial discretization of the Navier-Stokes equations is based on the nonconforming linear finite elements $\mathbf{P}^1_{nc}$ for the fluid velocity. Hence, we study in details nonconforming finite elements. In Section 7.4, we study $\mathbf{P}^{k+1}_{nc} \times P^k$ **discretization in the light of the discrete T-coercivity**. To do so, we need to exhibit a **Fortin operator**, that is specified for $k = 1$ $d = 2$ or 3 in §7.4.1, and for $k = 2$, $d = 2$ in §7.4.2, cf. [A9, Sections 3.1 and 3.2]. We give estimates of the stability constants in both cases. In §7.4.3, we illustrate the importance of using a divergence-free velocity on some numerical experiments, cf. [A9, §4].

In order to improve the accuracy of the $\mathbf{P}^1_{nc} \times P^0$ scheme, one can enrich the space of the discrete pressure, as it is done in the TrioCFD code, where the discrete pressure has a P^0 component plus a P^1-Lagrange component. This leads to the so called $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ scheme, which exhibits interesting numerical features. However, only an incomplete proof of the stability is available in [175, 153]. In Section 7.5, **we carry out the numerical analysis of the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ scheme**, and we provide some numerical illustrations cf. [P7, W1].

My research on nonconforming finite elements led me to supervise Andrew Peitavy's PhD thesis [210]. In order to improve the accuracy of the $\mathbf{P}^1_{nc} \times P^0$ scheme without increasing the number of degrees of

freedom, we can use **the symmetric Multipoints Flux Approximation scheme to discretize the pressure gradient**, cf. proceedings [P3, P4]. We give details of this discretization in Section 7.6.

In Section 7.7, **using the T-coercivity theory, we propose a new variational formulation of the Stokes problem** which doesn't involve nonlocal operators. With this new formulation, **unstable finite element pairs are stabilized**. In addition, the numerical scheme is easy to implement, and a better approximation of the velocity and the pressure is observed numerically when the viscosity is small cf. [A5].

Finally, in Chapter 8, we propose some extensions of this work.

Part II

Contributions

Chapter 4

Mathematical tools

In this Chapter, we introduce some useful mathematical tools. Throughout this manuscript, C is used to denote a generic positive constant which is independent of the mesh size, the triangulation and the quantities/fields of interest. We also use the shorthand notation $A \lesssim B$ for the inequality $A \leq CB$, where A and B are two scalar quantities, and C is a generic constant. Respectively, $A \eqsim B$ for the inequalities $A \lesssim B$ and $B \lesssim A$.

Vector-valued (resp. tensor-valued) function spaces are written in boldface character (resp. blackboard bold characters).

We call $\mathcal{O}$ a domain of $\mathbb{R}^d$ $(d = 1, 2, 3)$ iff $\mathcal{O}$ is a subset of $\mathbb{R}^d$, open, bounded, connected, with a Lipschitz boundary $\partial\mathcal{O}$. For the ease of exposition and as far as discretization is concerned, we assume that $\mathcal{O}$ is a Lipschitz polyhedron. We denote by $\mathbf{n}$ the unit normal vector from $\partial\mathcal{O}$. We generically consider $\Gamma \subset \partial\mathcal{O}$ a subset of $\partial\mathcal{O}$ of non-vanishing measure, with Lipschitz boundary $\partial\Gamma$. Finally, $C^\infty(\overline{\mathcal{O}})$ is the restriction to $\overline{\mathcal{O}}$ of $C^\infty(\mathbb{R}^d)$ functions.

In Section 4.1, we define the main functional spaces that we need. We have to treat boundary conditions or interface conditions, so we give the trace theorems that we will use in Section 4.2. In Section 4.3, we remind that there exists a (nonlocal) right inverse of the divergence operator. We provide regularity properties of solutions to elliptic problems in case of heterogeneous material and or non convex domain in Section 4.4. In Section 4.5, we summarize numerical analysis tools for mixed problems. We recall T-coercivity theory as written in [92, 85] in the case of complex-valued functions in Section 4.6. Finally, in Section 4.7, we define some common notations for the discretization.

4.1 Continuous settings

We recall that the functional space $L^p(\mathcal{O})$ contains classes of functions f measurable in the Lebesgue sense on $\mathcal{O}$, and such that

$$\begin{cases} \text{if } 1 \leq p < \infty : & \|f\|_{L^p(\mathcal{O})} := \left\{ \displaystyle\int_{\mathcal{O}} |f|^p \right\}^{1/p} < \infty \\ \text{if } p = \infty : & \|f\|_{L^\infty(\mathcal{O})} := \operatorname{esssup}_{\mathbf{x}\in\mathcal{O}} |f(\mathbf{x})| < \infty \end{cases} .$$

These functions are assumed to be real-valued[1]. For $p \in [1, \infty]$, $L^p(\mathcal{O})$ is a Banach space endowed with the norm $\|\cdot\|_{L^p(\mathcal{O})}$. And, for $1 \leq p < \infty$, $L^p(\mathcal{O})$ is separable. Finally, for $p \in [1, \infty]$, $W^{1,p}(\mathcal{O}) := \{v \in L^p(\mathcal{O}) \ : \ \partial_i v \in L^p(\mathcal{O}),\ 1 \leq i \leq d\}$ is a Banach space endowed with the product norm.

We define the following functional spaces on $\mathcal{O}$, which are Hilbert spaces endowed with their associated scalar product:

[1] In case they're complex-valued, we take into account the conjugation.

- $L^2(\mathcal{O})$, $(u,v)_{L^2(\mathcal{O})} = \displaystyle\int_{\mathcal{O}} u\,v$, $L^2_{zmv}(\mathcal{O}) := \left\{ u \in L^2(\mathcal{O}) \,|\, \displaystyle\int_{\mathcal{O}} u = 0 \right\}$;
  $\mathbf{L}^2(\mathcal{O}) := L^2(\mathcal{O})^d$, $(\mathbf{p},\mathbf{q})_{\mathbf{L}^2(\mathcal{O})} = \displaystyle\int_{\mathcal{O}} \mathbf{p}\cdot\mathbf{q}$.
- $H^1(\mathcal{O}) := \left\{ v \in L^2(\mathcal{O}) \ : \ \partial_i v \in L^2(\mathcal{O}),\ 1 \le i \le d \right\} = \left\{ v \in L^2(\mathcal{O}) \ : \ \nabla v \in \mathbf{L}^2(\mathcal{O}) \right\}$,
  $(u,v)_{H^1(\mathcal{O})} = (u,v)_{L^2(\mathcal{O})} + \displaystyle\sum_{i=1}^{d} (\partial_i u, \partial_i v)_{L^2(\mathcal{O})} = (u,v)_{L^2(\mathcal{O})} + (\nabla u, \nabla v)_{\mathbf{L}^2(\mathcal{O})}$,
  for $m \ge 2$: $H^m(\mathcal{O}) := \left\{ v \in L^2(\mathcal{O}) \ : \ \partial_i v \in H^{m-1}(\mathcal{O}),\ 1 \le i \le d \right\}$,
  $(u,v)_{H^m(\mathcal{O})} = (u,v)_{L^2(\mathcal{O})} + \displaystyle\sum_{i=1}^{d} (\partial_i u, \partial_i v)_{H^{m-1}(\mathcal{O})}$,
  for $s \in \mathbb{R}^+ \setminus \mathbb{N}$: we let $s = m + \sigma$, $m \in \mathbb{N}$ and $\sigma \in]0,1[$ and we define $H^s(\mathcal{O})$ by interpolation between $H^m(\mathcal{O})$ and $H^{m+1}(\mathcal{O})$, with the convention $H^0(\mathcal{O}) = L^2(\mathcal{O})$;
  for $s \ge 0$: $H^s_0(\mathcal{O})$ is defined as the closure of $\mathcal{D}(\mathcal{O})$ in $H^s(\mathcal{O})$;
  for $s \ge 0$, $C^\infty(\overline{\mathcal{O}})$ is dense in $H^s(\mathcal{O})$.
- $\mathbf{H}(\mathrm{div},\mathcal{O}) := \left\{ \mathbf{q} \in \mathbf{L}^2(\mathcal{O}) \ : \ \mathrm{div}\,\mathbf{q} \in L^2(\mathcal{O}) \right\}$, $(\mathbf{p},\mathbf{q})_{\mathbf{H}(\mathrm{div},\mathcal{O})} = (\mathbf{p},\mathbf{q})_{\mathbf{L}^2(\mathcal{O})} + (\mathrm{div}\,\mathbf{p}, \mathrm{div}\,\mathbf{q})_{L^2(\mathcal{O})}$;
  $\mathbf{H}_0(\mathrm{div},\mathcal{O})$ is defined as the closure of $\mathcal{D}(\mathcal{O})$ in $\mathbf{H}(\mathrm{div},\mathcal{O})$; $C^\infty(\overline{\mathcal{O}})^d$ is dense in $\mathbf{H}(\mathrm{div},\mathcal{O})$.
- $\mathbf{H}(\mathbf{curl},\mathcal{O}) := \left\{ \mathbf{q} \in \mathbf{L}^2(\mathcal{O}) \ : \ \mathbf{curl}\,\mathbf{q} \in \mathbf{L}^2(\mathcal{O}) \right\}$, $(\mathbf{p},\mathbf{q})_{\mathbf{H}(\mathbf{curl},\mathcal{O})} = (\mathbf{p},\mathbf{q})_{\mathbf{L}^2(\mathcal{O})} + (\mathbf{curl}\,\mathbf{p}, \mathbf{curl}\,\mathbf{q})_{\mathbf{L}^2(\mathcal{O})}$;
  $\mathbf{H}_0(\mathbf{curl},\mathcal{O})$ is defined as the closure of $\mathcal{D}(\mathcal{O})$ in $\mathbf{H}(\mathbf{curl},\mathcal{O})$;
  $C^\infty(\overline{\mathcal{O}})^d$ is dense in $\mathbf{H}(\mathbf{curl},\mathcal{O})$.

As $\mathcal{O}$ is a domain of $\mathbb{R}^d$, we can define its unit outward normal $\mathbf{n}$ a.e. on the boudary $\partial\mathcal{O}$, and we have $\mathbf{n} \in \mathbf{L}^\infty(\partial\mathcal{O})$. For a part $\Gamma \subset \partial\mathcal{O}$, we define the following functional spaces, which are Hilbert spaces endowed with their associated scalar product:

- $L^2(\Gamma) := \left\{ \mu \text{ measurable on } \Gamma \ : \ \displaystyle\int_\Gamma |\mu|^2 < +\infty \right\}$, $(\mu,\mu')_{L^2(\Gamma)} = \displaystyle\int_\Gamma \mu\,\mu'$.
- $H^{\frac12}(\Gamma) = \left\{ \mu \in L^2(\Gamma) \ : \ \displaystyle\int_{\Gamma_{\mathbf{x}}} \int_{\Gamma_{\mathbf{y}}} \frac{|\mu(\mathbf{x}) - \mu(\mathbf{y})|^2}{|\mathbf{x}-\mathbf{y}|^d} < \infty \right\}$,
  $(\mu,\mu')_{H^{1/2}(\Gamma)} = (\mu,\mu')_{L^2(\Gamma)} + \displaystyle\int_{\Gamma_{\mathbf{x}}} \int_{\Gamma_{\mathbf{y}}} \frac{(\mu(\mathbf{x}) - \mu(\mathbf{y}))\,(\mu'(\mathbf{x}) - \mu'(\mathbf{y}))}{|\mathbf{x}-\mathbf{y}|^d}$.
- For $s = 1$ or 3, we will need $\widetilde{H}^{\frac{s}{2}}(\Gamma) = \left\{ u \in H^{\frac{s}{2}}(\Gamma) \ : \ \tilde{u} \in H^{\frac{s}{2}}(\partial\mathcal{O}) \right\}$, where $\tilde{u} \in L^2(\partial\mathcal{O})$ is the extension of u on all $\partial\mathcal{O}$, equal to 0 on $\partial\mathcal{O} \setminus \bar{\Gamma}$.

The following strict inclusions hold: $H^{\frac12}(\Gamma) \subsetneq L^2(\Gamma) \subsetneq (H^{\frac12}(\Gamma))'$ and $\widetilde{H}^{\frac12}(\Gamma) \subsetneq L^2(\Gamma) \subsetneq (\widetilde{H}^{\frac12}(\Gamma))'$. Notice that if $\lambda \in L^2(\Gamma)$ and $\mu \in H^{\frac12}(\Gamma)$, it holds: $\langle \lambda, \mu \rangle_{H^{1/2}(\Gamma)} = \int_\Gamma \lambda\,\mu$.

4.2 Trace theorems

For $u \in C^\infty(\overline{\mathcal{O}})$ we denote by $u_{|\Gamma}$ its trace on a subset $\Gamma \subset \partial\mathcal{O}$.

4.2.1 Traces of $H^1(\mathcal{O})$ fonctions

Theorem 4.2.1. *The trace mapping $\gamma_0 : u \mapsto u_{|\partial\mathcal{O}}$ from $C^\infty(\overline{\mathcal{O}})$ on $\partial\mathcal{O}$ extends by continuity into a continuous mapping from $H^1(\mathcal{O})$ onto $L^2(\partial\mathcal{O})$: $\exists c > 0,\ \forall u \in H^1(\mathcal{O}), \quad \|\gamma_0(u)\|_{L^2(\partial\mathcal{O})} \le c\|u\|_{H^1(\mathcal{O})}$. The trace mapping is surjective and continuous from $H^1(\mathcal{O})$ onto $H^{\frac12}(\partial\mathcal{O})$:*

$$\exists c > 0,\ \forall u \in H^1(\mathcal{O}), \quad \|\gamma_0(u)\|_{H^{1/2}(\partial\mathcal{O})} \le c\|u\|_{H^1(\mathcal{O})};$$
$$\exists c > 0,\ \forall \mu \in H^{\frac12}(\partial\mathcal{O}), \quad \exists u \in H^1(\mathcal{O}) \text{ s.t. } \gamma_0(u) = \mu \text{ and } \|u\|_{H^1(\mathcal{O})} \le c\|\mu\|_{H^{1/2}(\partial\mathcal{O})},$$

where the $u \in H^1(\mathcal{O})$ is called a lifting of $\mu \in H^{\frac{1}{2}}(\partial\mathcal{O})$.

We have the identification: $H_0^1(\mathcal{O}) = \left\{v \in H^1(\mathcal{O}) \ : \ v_{|\partial\mathcal{O}} = 0\right\}$.

We can also define a trace mapping on Γ, from $H^1(\mathcal{O})$ onto $L^2(\Gamma)$, where the trace is denoted by $v_{|\Gamma}$. The trace mapping is surjective from $H^1(\mathcal{O})$ onto $H^{\frac{1}{2}}(\Gamma)$. Let $C_\Gamma^\infty(\overline{\mathcal{O}}) := \{v \in C^\infty(\overline{\mathcal{O}}) : v = 0$ in a neighbourhood of $\Gamma\}$. Calling $H_{0,\Gamma}^1(\mathcal{O})$ the closure of $C_\Gamma^\infty(\overline{\mathcal{O}})$ in $H^1(\mathcal{O})$, we can prove that: $H_{0,\Gamma}^1(\mathcal{O}) = \left\{v \in H^1(\mathcal{O}) \ : \ v_{|\Gamma} = 0\right\}$. In addition, the space of all traces on Γ of $C^\infty_{\partial\mathcal{O}\setminus\bar{\Gamma}}(\overline{\mathcal{O}})$ is dense in $L^2(\Gamma)$. Finally, note that if $v \in H^1(\mathcal{O})$ vanishes in a neighbourhood of $\partial\Gamma$, then $v_{|\Gamma} \in \widetilde{H}^{\frac{1}{2}}(\Gamma)$.

Theorem 4.2.2. (`Poincaré-Steklov inequality`) *Let $\mathcal{O}$ be a domain in $\mathbb{R}^d$ and $\Gamma \subset \partial\mathcal{O}$ a non empty subset. Then*

$$\exists C_{PS} > 0, \ \forall v \in H_{0,\Gamma}^1(\mathcal{O}), \quad \|v\|_{L^2(\mathcal{O})} \leq C_{PS}\, h_{\mathcal{O}} \|\nabla v\|_{\mathbf{L}^2(\mathcal{O})}.$$

Thanks to Theorem 4.2.2, in $H_0^1(\mathcal{O})$, the semi-norm is equivalent to the natural norm, so that the scalar product reads $(v,w)_{H_0^1(\mathcal{O})} = (\nabla v, \nabla w)_{\mathbf{L}^2(\mathcal{O})}$ and the norm is $\|v\|_{H_0^1(\mathcal{O})} = \|\nabla v\|_{\mathbf{L}^2(\mathcal{O})}$. Let $\mathbf{v}, \mathbf{w} \in \mathbf{H}_0^1(\mathcal{O})$. We denote by $(v_i)_{i=1}^d$ (resp. $(w_i)_{i=1}^d$) the components of $\mathbf{v}$ (resp. $\mathbf{w}$), and we set $\boldsymbol{\nabla}\,\mathbf{v} = (\partial_j v_i)_{i,j=1}^d \in \mathbb{L}^2(\mathcal{O})$ (where $\mathbb{L}^2(\mathcal{O}) = [L^2(\mathcal{O})]^{d\times d}$). We have:

$$(\boldsymbol{\nabla}\,\mathbf{v}, \boldsymbol{\nabla}\,\mathbf{w})_{\mathbb{L}^2(\mathcal{O})} = (\mathbf{v},\mathbf{w})_{\mathbf{H}_0^1(\mathcal{O})} = \sum_{i=1}^d (v_i, w_i)_{H_0^1(\mathcal{O})} \text{ and } \|\mathbf{v}\|_{\mathbf{H}_0^1(\mathcal{O})} = \left(\sum_{i=1}^d \|v_i\|^2_{H_0^1(\mathcal{O})}\right)^{\frac{1}{2}} = \|\boldsymbol{\nabla}\,\mathbf{v}\|_{\mathbb{L}^2(\mathcal{O})}.$$

4.2.2 Normal traces of $\mathbf{H}(\mathrm{div},\mathcal{O})$ functions

Theorem 4.2.3. *The normal trace mapping $\gamma_1 : \mathbf{q} \mapsto \gamma_1(\mathbf{q}) = \mathbf{q}\cdot\mathbf{n}_{|\partial\mathcal{O}}$ from $\mathcal{C}^\infty(\overline{\mathcal{O}})^d$ onto $\partial\mathcal{O}$ extends by continuity into a continuous and surjective application of $\mathbf{H}(\mathrm{div},\mathcal{O})$ in the dual space $(H^{\frac{1}{2}}(\partial\mathcal{O}))'$:*

$$\exists c > 0, \ \forall \mathbf{q} \in \mathbf{H}(\mathrm{div},\mathcal{O}), \quad \|\gamma_1(\mathbf{q})\|_{(H^{\frac{1}{2}}(\partial\mathcal{O}))'} \leq c\|\mathbf{q}\|_{\mathbf{H}(\mathrm{div},\mathcal{O})} ;$$
$$\exists c > 0, \ \forall \mu \in (H^{\frac{1}{2}}(\partial\mathcal{O}))', \ \exists \mathbf{q} \in \mathbf{H}(\mathrm{div},\mathcal{O}) \ s.t. \ \gamma_1(\mathbf{q}) = \mu \ and \ \|\mathbf{q}\|_{\mathbf{H}(\mathrm{div},\mathcal{O})} \leq c\|\mu\|_{(H^{\frac{1}{2}}(\partial\mathcal{O}))'}.$$

For the last point, the vector $\mathbf{q} \in \mathbf{H}(\mathrm{div},\mathcal{O})$ is called a lifting of $g \in (H^{\frac{1}{2}}(\partial\mathcal{O}))'$.
We have the identification:

$$\mathbf{H}_0(\mathrm{div},\mathcal{O}) = \left\{\mathbf{q} \in \mathbf{H}(\mathrm{div},\mathcal{O}) \ : \ \mathbf{q}\cdot\mathbf{n}_{|\partial\mathcal{O}} = 0\right\}.$$

Let Γ be a strict subset of $\partial\mathcal{O}$. For $\mathbf{q} \in \mathbf{H}(\mathrm{div},\mathcal{O})$, we know from Theorem 4.2.3 that $\mathbf{q}\cdot\mathbf{n}_{|\partial\mathcal{O}} \in (H^{\frac{1}{2}}(\partial\mathcal{O}))'$. However, in general $\mathbf{q}\cdot\mathbf{n}_{|\Gamma} \notin (H^{\frac{1}{2}}(\Gamma))'$, while we automatically have $\mathbf{q}\cdot\mathbf{n}_{|\Gamma} \in (\widetilde{H}^{\frac{1}{2}}(\Gamma))'$.

4.2.3 Tangential traces of $\mathbf{H}(\mathbf{curl},\mathcal{O})$ functions

Theorem 4.2.4. *The tangential trace mapping $\gamma_T : \mathbf{q} \mapsto \gamma_T(\mathbf{q}) = \mathbf{q}\times\mathbf{n}_{|\partial\mathcal{O}}$ from $\mathcal{C}^\infty(\overline{\mathcal{O}})^d$ onto $\partial\mathcal{O}$ extends by continuity into a continuous and surjective application of $\mathbf{H}(\mathbf{curl},\mathcal{O})$ in the dual space $(\mathbf{H}^{\frac{1}{2}}(\partial\mathcal{O}))'$. In addition, the following characterization holds:*

$$\mathbf{H}_0(\mathbf{curl},\mathcal{O}) = \left\{\mathbf{q} \in \mathbf{H}(\mathbf{curl},\mathcal{O}) \ : \ \mathbf{q}\times\mathbf{n}_{|\partial\mathcal{O}} = 0\right\}.$$

The space of the tangential traces of $\mathbf{H}(\mathbf{curl},\mathcal{O})$ functions is characterized in [74].

4.2.4 Integration by parts formulas

The following integration by parts formulas hold:

$$\forall v \in H^1_0(\mathcal{O}),\ \forall \mathbf{q} \in \mathbf{H}(\mathrm{div},\mathcal{O}),\ \int_{\mathcal{O}} (\mathbf{q}\cdot\nabla v + \mathrm{div}\,\mathbf{q}\, v) = 0, \tag{4.2.1}$$

$$\forall v \in H^1(\mathcal{O}),\ \forall \mathbf{q} \in \mathbf{H}(\mathrm{div},\mathcal{O}),\ \int_{\mathcal{O}} (\mathbf{q}\cdot\nabla v + \mathrm{div}\,\mathbf{q}\, v) = \langle \mathbf{q}\cdot\mathbf{n}\,,\, v\rangle_{H^{1/2}(\partial\mathcal{O})}. \tag{4.2.2}$$

If you want to "cut" the duality brackets in (4.2.2) on Γ and $\Gamma' = \partial\mathcal{O}\setminus\bar{\Gamma}$ two (strict) parts of $\partial\mathcal{O}$, there are two possibilities.
(4.2.2-i) Either assume that the normal trace $\mathbf{q}\cdot\mathbf{n}_{|\partial\mathcal{O}}$ is more regular than $(H^{\frac{1}{2}}(\partial\mathcal{O}))'$, for instance assume that $\mathbf{q}\cdot\mathbf{n}_{|\partial\mathcal{O}} \in L^2(\partial\mathcal{O})$. In that case, one can write that for all $v \in H^1(\mathcal{O})$:

$$\langle \mathbf{q}\cdot\mathbf{n}\,,\, v\rangle_{H^{1/2}(\partial\mathcal{O})} = \int_{\partial\mathcal{O}} \mathbf{q}\cdot\mathbf{n}\, v = \int_{\Gamma} \mathbf{q}\cdot\mathbf{n}\, v + \int_{\Gamma'} \mathbf{q}\cdot\mathbf{n}\, v.$$

(4.2.2-ii) Either you must choose $v \in H^1(\mathcal{O})$ vanishing in the neighbourhood of $\partial\Gamma = \partial\Gamma'$. In that case, $v_{|\Gamma} \in \widetilde{H}^{\frac{1}{2}}(\Gamma)$ and $v_{|\Gamma'} \in \widetilde{H}^{\frac{1}{2}}(\Gamma')$, and we can write that for all $\mathbf{q} \in \mathbf{H}(\mathrm{div},\mathcal{O})$:

$$\langle \mathbf{q}\cdot\mathbf{n}\,,\, v\rangle_{H^{1/2}(\partial\mathcal{O})} = \langle \mathbf{q}\cdot\mathbf{n}_{|\Gamma}\,,\, v_{|\Gamma}\rangle_{\widetilde{H}^{1/2}(\Gamma)} + \langle \mathbf{q}\cdot\mathbf{n}_{|\Gamma'}\,,\, v_{|\Gamma'}\rangle_{\widetilde{H}^{1/2}(\Gamma')}.$$

We also recall the following integration formula:

$$\forall \mathbf{v}, \in \mathbf{H}(\mathbf{curl},\mathcal{O}),\ \forall\, \mathbf{w} \in \mathbf{H}_0(\mathbf{curl},\mathcal{O}),\ \int_{\mathcal{O}} (\mathbf{curl}\,\mathbf{v}\cdot\mathbf{w} - \mathbf{curl}\,\mathbf{w}\cdot\mathbf{v}) = 0. \tag{4.2.3}$$

4.3 Babuška-Aziz constant

In Chapter 7, we need to prove the well-posedness of Stokes problem, which rely on the proposition below. Let Ω be the domain of study. Let us set $\mathbf{V} = \big\{\mathbf{v} \in \mathbf{H}^1_0(\Omega) \mid \mathrm{div}\,\mathbf{v} = 0\big\}$. The vector space $\mathbf{V}$ is a closed subset of $\mathbf{H}^1_0(\Omega)$. We denote by $\mathbf{V}^\perp$ the orthogonal of $\mathbf{V}$ in $\mathbf{H}^1_0(\Omega)$. We have the

Proposition 4.3.1 (`[8, Theorem 3.1 (e)]`). *The operator* $\mathrm{div} : \mathbf{H}^1_0(\Omega) \to L^2(\Omega)$ *is an isomorphism of* $\mathbf{V}^\perp$ *onto* $L^2_{zmv}(\Omega)$. *We call* C_{div} *the constant such that:*

$$\forall p \in L^2_{zmv}(\Omega),\ \exists! \mathbf{v} \in \mathbf{V}^\perp \mid \mathrm{div}\,\mathbf{v} = p \text{ and } \|\mathbf{v}\|_{\mathbf{H}^1_0(\Omega)} \le C_{\mathrm{div}}\|p\|_{L^2(\Omega)}. \tag{4.3.1}$$

The constant C_{div} depends only on the domain Ω. Recall that we have: $\|\mathbf{v}\|^2_{\mathbf{H}^1_0(\Omega)} = \|\,\mathbf{curl}\,\mathbf{v}\|^2_{\mathbf{L}^2(\Omega)} + \|\,\mathrm{div}\,\mathbf{v}\|^2_{L^2(\Omega)} \ge \|p\|^2_{L^2(\Omega)}$, hence $C_{\mathrm{div}} \ge 1$. Using proposition 4.3.1, we can define a lifting in $\mathbf{H}^1(\Omega)$ of a function $v \in L^2(\Omega)$:

Corollary 4.3.1 (`[A9, Corollary 4]`). *For all* $v \in L^2(\Omega)$, *it exists* $\mathbf{s} \in \mathbf{H}^1(\Omega)$ *such that:*

$$\mathrm{div}\,\mathbf{s} = v \text{ and } \|\mathbf{s}\|_{\mathbf{L}^2(\Omega)} + h_\Omega\|\,\boldsymbol{\nabla}\,\mathbf{s}\|_{\mathbb{L}^2(\Omega)} \le C_\Omega\, h_\Omega\|v\|_{L^2(\Omega)}, \tag{4.3.2}$$

where the dimensionless constant C_Ω *depends on* C_{div} *and* C_{PS}.

Actually, the constant C_{div} is such that $C_{\mathrm{div}} = 1/\beta(\Omega)$ where $\beta(\Omega)$, known as the Babuška-Aziz constant, is the inf-sup condition (or Ladyzhenskaya–Babuška–Brezzi condition):

$$\beta(\Omega) = \inf_{q\in L^2_{zmv}(\Omega)\setminus\{0\}} \sup_{\mathbf{v}\in\mathbf{H}^1_0(\Omega)\setminus\{0\}} \frac{(q, \mathrm{div}\,\mathbf{v})_{L^2(\Omega)}}{\|q\|_{L^2(\Omega)}\,\|\mathbf{v}\|_{\mathbf{H}^1_0(\Omega)}}. \tag{4.3.3}$$

Generally, the value of $\beta(\Omega)$ is not known explicitly. In [41], Bernardi et al established results on the

discrete approximation of $\beta(\Omega)$ using conforming finite elements. Recently, Gallistl proposed in [154] a numerical scheme with adaptive meshes for computing approximations to $\beta(\Omega)$. In the case of $d = 2$, Costabel and Dauge [118] established the following bound:

Theorem 4.3.1. *Let $\Omega \subset \mathbb{R}^2$ be a domain contained in a ball of radius R, star-shaped with respect to a concentric ball of radius ρ. Then*

$$\beta(\Omega) \geq \frac{\rho}{\sqrt{2}\,R}\left(1+\sqrt{1-\frac{\rho^2}{R^2}}\right)^{-\frac{1}{2}} \geq \frac{\rho}{2\,R}. \tag{4.3.4}$$

Let us detail the bound for some remarkable domains. If Ω is a ball, $\beta(\Omega) \geq \frac{1}{2}$ and if Ω is a square, $\beta(\Omega) \geq \frac{1}{2\sqrt{2}}$. Suppose now that Ω is stretched in some direction by a factor k, then $\beta(\Omega) \geq \frac{1}{2k}$. Finally, if Ω is L-shaped (resp. cross-shaped) such that $L = k\,l$, where L is the largest length and l is the smallest length of an edge, then $\beta(\Omega) \geq \frac{1}{2\sqrt{2}\,k}$ (resp. $\beta(\Omega) \geq \frac{1}{4\,k}$).

4.4 Regularity of elliptic solutions

4.4.1 Heterogeneous materials

In Chapter 5, we consider magnetostatic systems in heterogeneous material (Sections 5.2 and 5.4). In Chapter 6, our numerical results concern the simulation of pressurized water reactor cores, which are also modelled by heterogeneous material. In both cases, the numerical analysis must be carried out with care, taking into account the low-regularity of the solutions.

We split the domain $\mathcal{R} \subset \mathbb{R}^d$ into $N_{\mathcal{R}}$ open disjoint parts $\{\mathcal{R}_i\}_{i\in\mathcal{I}_{\mathcal{R}}}$, where $\mathcal{I}_{\mathcal{R}} = \{1,\ldots,N_{\mathcal{R}}\}$, with Lipschitz, piecewise smooth boundaries: $\overline{\mathcal{R}} = \cup_{i\in\mathcal{I}_{\mathcal{R}}}\overline{\mathcal{R}_i}$ and the set $\{\mathcal{R}_i\}_{i\in\mathcal{I}_{\mathcal{R}}}$ is called a partition of $\mathcal{R}$. For a field v defined over $\mathcal{R}$, we shall use the notations $v_i = v_{|\mathcal{R}_i}$, for $i \in \mathcal{I}_{\mathcal{R}}$.
Given a partition $\{\mathcal{R}_i\}_{i\in\mathcal{I}_{\mathcal{R}}}$ of $\mathcal{R}$, we introduce function spaces with piecewise regular elements:

$$\begin{aligned}
\mathcal{P}H^s(\mathcal{R}) &= \left\{\psi \in L^2(\mathcal{R}) \,|\, \psi_i \in H^s(\mathcal{R}_i),\, i \in \mathcal{I}_{\mathcal{R}}\right\},\ s > 0\,;\\
\mathcal{P}\mathbf{H}(\mathrm{div},\mathcal{R}) &= \left\{\mathbf{q} \in \mathbf{L}^2(\mathcal{R}) \,|\, \mathbf{q}_i \in \mathbf{H}(\mathrm{div},\mathcal{R}_i),\, i \in \mathcal{I}_{\mathcal{R}}\right\};\\
\mathcal{P}W^{1,\infty}(\mathcal{R}) &= \left\{\psi \in L^\infty(\mathcal{R}) \,|\, \psi_i \in W^{1,\infty}(\mathcal{R}_i),\, i \in \mathcal{I}_{\mathcal{R}}\right\}.
\end{aligned}$$

We recall that for a piecewise smooth $\psi \in \mathcal{P}H^s(\mathcal{R})$, $\|\psi\|^2_{\mathcal{P}H^s(\mathcal{R})} = \sum_{i\in\mathcal{I}_{\mathcal{R}}} \|\psi\|^2_{H^s(\mathcal{R}_i)}$. Similarly for elements of $\mathcal{P}W^{1,\infty}(\mathcal{R})$. Let $\mathcal{R} \subset \mathbb{R}^d$, $d \in \{2,3\}$. We split the boundary into three disjoint, open parts such that: $\partial\mathcal{R} = \overline{\Gamma}_D \cup \overline{\Gamma}_N \cup \overline{\Gamma}_R$. Consider the generalized Laplacian problem:

$$\left\{\begin{array}{l}
\text{Find } u \text{ such that}\\
\quad -\,\mathrm{div}\,\mathbb{k}\,\nabla u + q\,u = f \text{ in } \mathcal{R},\\
\quad u_{|\Gamma_D} = g_D \text{ on } \Gamma_D,\\
\quad \nabla u\cdot\mathbf{n}_{|\Gamma_N} = g_N \text{ on } \Gamma_N,\\
\quad \nabla u\cdot\mathbf{n}_{|\Gamma_R} + \alpha u_{|\Gamma_R} = g_R \text{ on } \Gamma_R.
\end{array}\right. \tag{4.4.1}$$

We suppose that $g_D \in \widetilde{H}^{\frac{1}{2}}(\Gamma_D)$, $g_N \in L^2(\Gamma_N)$ and $g_R \in L^2(\Gamma_R)$.
The coefficients q, respectively $\mathbb{k}$, are a scalar field, resp. a tensor field, such that:

$$\left\{\begin{array}{l}
(q,\mathbb{k}) \in L^\infty(\mathcal{R}) \times \mathbb{L}^\infty(\mathcal{R}),\\
\exists \kappa_*,\kappa^* > 0,\ \forall \mathbf{z} \in \mathbb{R}^d,\ \kappa_*\|\mathbf{z}\|^2 \leq (\mathbb{k}\mathbf{z},\mathbf{z}) \leq \kappa^*\|\mathbf{z}\|^2 \text{ a.e. in } \mathcal{R},\\
\exists q_* \geq 0, q^* > 0,\ q_* \leq q \leq q^* \text{ a.e. in } \mathcal{R}.
\end{array}\right. \tag{4.4.2}$$

Proposition 1 of [A6], which is written in the case where $\Gamma_D = \partial\mathcal{O}$, $g_D = 0$, holds also for solution to Problem (4.4.1).

Proposition 4.4.1 ([A6, Proposition 1]). *Let $q \in \mathcal{P}W^{1,\infty}(\mathcal{R})$, $\mathbb{k} \in \mathcal{P}\mathbb{W}^{1,\infty}(\mathcal{R})$ a symmetric tensor field, that fulfill (4.4.2). There exists $r_{\max} \in]0,1]$, called the regularity exponent, such that for all source terms $f \in L^2(\mathcal{R})$, the solution $u \in H^1(\mathcal{R})$ to problem (4.4.1) belongs to the space $\mathcal{P}\dot{H}^{1+r_{\max}}(\mathcal{R})$ such that $\mathcal{P}\dot{H}^{1+r_{\max}}(\mathcal{R}) := \bigcap_{0\le r<r_{\max}} \mathcal{P}H^{1+r}(\mathcal{R})$ ($r_{\max} < 1$) or $\mathcal{P}H^2(\mathcal{R})$ ($r_{\max} = 1$) with continuous dependence: for all $r \in [0, r_{\max}[$, $\|\phi\|_{\mathcal{P}H^{1+r}(\mathcal{R})} \lesssim_r \|f\|_{L^2(\mathcal{R})}$ ($r_{\max} < 1$) or $\|\phi\|_{\mathcal{P}H^2(\mathcal{R})} \lesssim \|f\|_{L^2(\mathcal{R})}$ ($r_{\max} = 1$).*

Proof. We refer to [119, Theorem 4.1], [55, Theorem 3.1] for homogeneous Dirichlet or Neuman boundary conditions. Otherwise, one considers a lifting $\tilde{u} \in H^1(\mathcal{R})$ such that:

$$\left\{\begin{array}{rcl} -\operatorname{div}\mathbb{k}\,\nabla\,\tilde{u} + q\,\tilde{u} &=& 0 \text{ in } \mathcal{R},\\ \tilde{u}_{|\Gamma_D} &=& g_D \text{ on } \Gamma_D,\\ \nabla\,\tilde{u}\cdot\mathbf{n}_{|\Gamma_N} &=& g_N \text{ on } \Gamma_N,\\ \nabla\,\tilde{u}\cdot\mathbf{n}_{|\Gamma_R} + \alpha\tilde{u}_{|\Gamma_R} &=& g_R \text{ on } \Gamma_R. \end{array}\right.$$

Following [179], on has: $\tilde{u} \in H^{\frac{3}{2}}(\mathcal{R})$. □

Since cross-points are allowed in our model, cf. Figure 6.1, and in accordance with [55], the low-regularity case corresponds precisely to $r_{\max} < 1/2$.

4.4.2 Non convex domains

In Chapter 5, we consider the electric eigenproblem and the wave propagation in non convex domains. In Chapter 7, we also need to consider that the domain of the fluid may be non convex. In both cases, once again, the numerical analysis must be carried out with care, taking into account the low-regularity of the solutions.

Let us now recall a classical shift theorem in the domain Ω. We call $\sigma_{Dir} \in]0,1]$ the limit regularity exponent for the Laplace problem with Dirichlet boundary condition

$$\text{Find } u \in H^1_0(\Omega) \text{ such that } -\Delta u = f \text{ in } \Omega$$

with data $f \in L^2(\Omega)$. According to [179], we know that $u \in H^{\frac{3}{2}}(\Omega)$, with $\|u\|_{H^{3/2}(\Omega)} \lesssim \|f\|_{L^2(\Omega)}$; hence $\sigma_{Dir} \ge 1/2$. In addition, cf. [166, 124]:

- if Ω is convex or if its boundary is of class $\mathcal{C}^2$, then $u \in H^2(\Omega)$, with $\|u\|_{H^2(\Omega)} \lesssim \|f\|_{L^2(\Omega)}$, and in this case $\sigma_{Dir} = 1$;
- if Ω is a non-convex polyhedron, then there exists $\sigma_{Dir} \in]1/2, 1[$ such that
  - $\forall f \in L^2(\Omega)$, $u \in \dot{H}^{1+\sigma_{Dir}}(\Omega) := \bigcap_{0\le s<\sigma_{Dir}} H^{1+s}(\Omega)$;
  - $\exists f \in L^2(\Omega)$ such that $u \notin H^{1+\sigma_{Dir}}(\Omega)$;
  - for each $s \in]0, \sigma_{Dir}[$, one has $\|u\|_{H^{1+s}(\Omega)} \lesssim_s \|f\|_{L^2(\Omega)}$.

We also recall the Birman-Solomyak decomposition, see [46, Theorem 4.1], also called the regular-gradient splitting, see in [176, Lemma 2.4]. Any element $\mathbf{u} \in \mathbf{H}_0(\mathbf{curl}, \Omega)$ can be split as

$$\begin{array}{l} \mathbf{u} = \mathbf{u}_{reg} + \nabla\, q_u \text{ in } \Omega, \text{ with } \mathbf{u}_{reg} \in \mathbf{H}^1(\Omega),\ q_u \in H^1_0(\Omega), \text{ and}\\ \|\mathbf{u}_{reg}\|_{\mathbf{H}^1(\Omega)} + \|q_u\|_{H^1(\Omega)} \lesssim \|\mathbf{u}\|_{\mathbf{H}(\mathbf{curl},\Omega)}. \end{array} \tag{4.4.3}$$

Then, if one changes the Dirichlet boundary condition to a Neumann boundary condition, one has similar results.

We refer to [115] for domains with cracks.

4.5 Discretization and convergence results for mixed problems

In Chapter 5, we will consider a mixed variational formulation for the electric eigenproblem, cf. Section 5.5 and for the propagation in a waveguide, cf. Section 5.6: the divergence constraint is taken into account with a Lagrange multiplier.

In Chapter 7, we will study Stokes problem, which is a also posed in a mixed variational formulation. The pressure of the fluid plays the role of a Lagrange multiplier on the divergence constraint.

4.5.1 Continuous mixed problems

We recall here the general theory (see [51]), which allows us to prove convergence results, for eigenvalue problems set in a mixed variational framework. So, consider

- V and Q two Hilbert spaces, with norms $\|\cdot\|_V$ and $\|\cdot\|_Q$;
- $a(\cdot,\cdot)$ a bilinear form, continuous, symmetric, positive, semidefinite form on $V \times V$;
- $b(\cdot,\cdot)$ a bilinear form, continuous form on $V \times Q$;
- (f, g) an element of $V' \times Q'$.

Introduce the abstract mixed variational problem:

$$\left\{ \begin{array}{l} \text{Find } (u,p) \in V \times Q \text{ such that} \\ a(u,v) + b(v,p) = \langle f, v\rangle \qquad \forall v \in V, \\ \qquad\quad b(v,p) = \langle g, q\rangle \qquad \forall q \in Q. \end{array} \right. \tag{4.5.1}$$

For this problem, we make two assumptions:

inf-sup condition: $\exists \beta > 0$ such that

$$\inf_{q\in Q\setminus\{0\}} \sup_{v\in V\setminus\{0\}} \frac{b(v,q)}{\|v\|_V \, \|q\|_Q} \geq \beta. \tag{4.5.2}$$

coercivity on the kernel: $\exists \alpha > 0$ such that

$$\forall v \in \mathbb{K}, \quad a(v,v) \geq \alpha \|v\|_V^2, \text{ where } \mathbb{K} = \{v \in V \,|\, \forall q \in Q,\, b(v,q) = 0\}. \tag{4.5.3}$$

Theorem 4.5.1 (`Babuška-Brezzi`, [161, 50])**.** *Assume that* (4.5.2) *and* (4.5.3) *are fulfilled. There exists one, and only one, solution* (u, p) *to the mixed variational problem* (4.5.1)*, which depends continuously on the data* (f, g)*.*

Note that if (4.5.2) and (4.5.3) hold, we can introduce the continuous and linear mapping `A` from V to V, such that `A`f is the first element (called u above) of the unique solution to problem (4.5.1), with data $(f, 0)$. Since a is symmetric, `A` is self-adjoint.

4.5.2 Conforming discretization

Then, one can consider the conforming discretization of (4.5.1). To that aim, let V_h and Q_h be two finite dimensional subspaces of V and Q respectively, and introduce the discrete mixed variational problem:

$$\left\{ \begin{array}{l} \text{Find } (u_h,p_h) \in V_h \times Q_h \text{ such that} \\ a(u_h,v_h) + b(v_h,p_h) = \langle f, v_h\rangle \quad \forall v_h \in V_h \quad . \\ \qquad\quad b(v_h,p_h) = \langle g, q_h\rangle \quad \forall q_h \in Q_h \end{array} \right. \tag{4.5.4}$$

For this discrete problem, we make two assumptions:

discrete inf-sup condition: $\forall h > 0$, $\exists \beta(h) > 0$ such that

$$\inf_{q_h \in Q \setminus \{0\}} \sup_{v_h \in V_h \setminus \{0\}} \frac{b(v_h, q_h)}{\|v_h\|_V \, \|q_h\|_Q} \geq \beta(h). \tag{4.5.5}$$

coercivity on the discrete kernel: $\exists \alpha > 0$ such that

$$\forall v \in \mathbb{K}_h, \quad a(v_h, v_h) \geq \alpha \|v_h\|_V^2, \text{ where } \mathbb{K}_h = \{v_h \in V_h \,|\, \forall q_h \in Q_h,\, b(v_h, q_h) = 0\}. \tag{4.5.6}$$

Theorem 4.5.2 (`Babuška-Brezzi`, [161, 50])**.** *Assume that* (4.5.2), (4.5.3), (4.5.5) *and* (4.5.6) *are fulfilled. There exists one, and only one, solution* (u_h, p_h) *to the discrete mixed variational problem* (4.5.4)*, with the error estimate below:*

$$\exists C > 0 \text{ such that: } \quad \|u - u_h\|_V \leq C \left\{ \left(1 + \frac{1}{\beta(h)} \inf_{v_h \in V_h} \|u - v_h\|_V + \inf_{q_h \in Q_h} \|p - q_h\|_Q \right) \right\}.$$

4.5.3 Resolution algorithm

Let $N_u = \dim(V_h)$, $N_p = \dim(Q_h)$, $\mathbb{A}_u \in \mathbb{R}^{N_u \times N_u}$ be the matrix related to the bilinear $a(\cdot,\cdot)$ and $\mathbb{B}_{p,u} \in \mathbb{R}^{N_p \times N_u}$ be the matrix related to the bilinear $b(\cdot,\cdot)$. Let $(\underline{U}, \underline{p}) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p}$ represent (u_h, p_h). Problem (4.5.4) corresponds to solving the following saddle point system:

$$\left\{ \begin{array}{l} \text{Find } (\underline{U}, \underline{P}) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p} \text{ such that} \\ \mathbb{A}_u \, \underline{U} + (\mathbb{B}_{p,u})^T \underline{P} \;=\; \underline{F} \\ \qquad\qquad \mathbb{B}_{p,u} \, \underline{U} \;=\; \underline{G} \end{array} \right. . \tag{4.5.7}$$

Let us set $\mathbb{K}_p = \mathbb{B}_{p,u} \, (\mathbb{A}_u)^{-1} (\mathbb{B}_{p,u})^T$. The matrix $\mathbb{K}_p \in \mathbb{R}^{N_p} \times \mathbb{R}^{N_p}$ is a symmetric matrix, furthermore it is positive definite as soon as the kernel of $(\mathbb{B}_{p,u})^T$ is reduced to $\{0\}$. When it is the case, the coupled problem (4.5.7) can be solved using the three + one steps below (the fourth step being straightforward):

Prediction:	Solve in $\underline{U}_\star$ such that $\mathbb{A}_u \, \underline{U}_\star = \underline{F}$.
Constraint solver:	Solve in $\underline{P}$ such that $\mathbb{K}_p \, \underline{P} = \mathbb{B}_{p,u} \, \underline{U}_\star - \underline{G}$.
Correction:	Solve in $\delta \underline{U}$ such that $\mathbb{A}_u \, \delta \underline{U} = -(\mathbb{B}_{p,u})^T \, \underline{P}$.
Update:	$\underline{U} = \delta \underline{U} + \underline{U}_\star$.

(4.5.8)

We can use the Cholesky factorization (computed once and for all) to solve linear systems with matrix $\mathbb{A}_u$. One can check easily that the above computed solution $(\underline{U}, \underline{P})$ solves (4.5.7). The constraint solver with matrix $\mathbb{K}_p$ is based on the Uzawa algorithm, which is the conjugate gradient algorithm in the context of the saddle point problem. It can be preconditioned by the inverse of the matrix $\mathbb{M}_p$ associated to the norm in Q_h (see e.g. [205, Lemma 5.9]). Thanks to the uniform discrete inf-sup condition, the number of iterations of the conjugate gradient algorithm is independent of the mesh size.

We will use this algorithm in Chapter 5, §5.6.3 to solve a problem of propagation in a waveguide; and Chapter 7, §7.7.2 to solve Stokes problem.

4.5.4 The first eigenproblem

Suppose that $g = 0$. Assume that (4.5.2), (4.5.3), (4.5.5) and (4.5.6) hold. For a given h, we can also introduce the discrete, continuous and linear mapping T_h from V to V , such that $\mathsf{T}_h f$ is the first element of the unique solution to problem (4.5.4), with data $(f, 0)$. We then consider the eigenproblem, set in a mixed variational framework. We need a third Hilbert space, called L, such that $V \subset L$ algebraically and topologically, and moreover V is a dense subset of L. The scalar product of L and its norm are

respectively denoted by $(\cdot,\cdot)_L$ and $\|\cdot\|_L$. Hereafter, we identify L' with L. The eigenproblem to be solved reads:

$$\left\{\begin{array}{l}\text{Find } (u,p,\lambda)\in V\times Q\times\mathbb{R} \text{ such that}\\ a(u,v)+b(v,p) = \lambda(u,v)_L \quad \forall v\in V,\\ b(u,q) = 0 \quad \forall q\in Q.\end{array}\right. \tag{4.5.9}$$

Note that, according to the definition of the operator T, problem (4.5.9) can be rewritten equivalently $\lambda\mathsf{T}u=u$, where T is considered from L to V. For this eigenproblem, we make one additional assumption:

$$\text{The operator } \mathsf{T}: L\to V \text{ is compact (and nonnegative).} \tag{4.5.10}$$

The discrete approximation of problem (4.5.9) is:
Find $(u_h,p_h,\lambda_h)\in V_h\times Q_h\times\mathbb{R}$ such that:

$$\left\{\begin{array}{rcll} a(u_h,v_h)+b(v_h,p_h) &=& \lambda_h(u_h,v_h)_L & \forall v_h\in V_h,\\ b(u_h,q_h) &=& 0 & \forall q_h\in Q_h.\end{array}\right. \tag{4.5.11}$$

The discrete eigenproblem can be rewritten equivalently $\lambda_h\mathsf{T}_h u_h=u_h$. So, to establish the convergence of the solution of the discrete eigenproblem towards the solution of the exact eigenproblem, one has to prove that

$$\lim_{h\to 0^+}\sup_{f\in L}\frac{\|(\mathsf{T}-\mathsf{T}_h)f\|_V}{\|f\|_L}=0. \tag{4.5.12}$$

Indeed, this uniform convergence in $\mathcal{L}(L,V)$ implies convergence of eigenvectors and eigenvalues, for T self-adjoint and compact [18]. Moreover, the (uniform) rate of convergence on the eigenvectors and eigenvalues is bounded by the rate of convergence in (4.5.12). From [51], we introduce sufficient conditions that ensure (4.5.12). A few additional definitions are needed, before these conditions are stated. Let $\mathcal{S}_0$ denote the subspace of $V\times Q$, made of solutions to problem (4.5.1) with right-hand sides $(f,0)$, where $f\in L$. Then, consider

$$V_0=\{v\in V:\ \exists q\in Q \text{ s.t. } (v,q)\in\mathcal{S}_0\},\ \text{and } Q_0=\{q\in Q:\ \exists v\in V \text{ s.t. } (v,q)\in\mathcal{S}_0\}. \tag{4.5.13}$$

These two spaces are endowed with their *natural* norms, that is $\|v\|_{V_0}$ (resp. $\|q\|_{Q_0}$) is equal to $\|f\|_L$ (resp. $\|f\|_L$), with f such that $(f,0)$ is the right-hand side of (4.5.1) with solution (v,q_f) (resp. solution (v_f,q)). The first condition is the *weak approximability* of Q_0.

Definition 4.5.1. The weak approximability of Q_0 is verified provided there exists $r_1:\mathbb{R}^+\to\mathbb{R}^+$, such that $\lim_{h\to 0^+} r_1(h)=0$ and

$$\sup_{v_h\in\mathbb{K}_h}\frac{b(v_h,q_0)}{\|v_h\|_V}\le r_1(h)\|q_0\|_{Q_0},\quad \forall q_0\in Q_0. \tag{4.5.14}$$

Inequality (4.5.14) actually corresponds to an appoximability property, since one has $b(v_h,q_h)=0$, $\forall q_h\in Q_h$ and $\forall v_h\in\mathbb{K}_h$.

The second condition is the *strong approximability* of V_0.

Definition 4.5.2. The strong approximability of V_0 is verified provided there exists $r_2:\mathbb{R}^+\to\mathbb{R}^+$, such that $\lim_{h\to 0^+} r_2(h)=0$ and

$$\forall v_0\in V_0,\quad \exists v_{0,\mathbb{K}_h}\in\mathbb{K}_h \text{ s.t. } \|v_0-v_{0,\mathbb{K}_h}\|_V\le r_2(h)\|v_0\|_{V_0}. \tag{4.5.15}$$

We note that according to Theorem 4.5.2, provided the weak approximability is fulfilled, it is sufficient that

$$\lim_{h\to 0^+}\left\{\frac{1}{\beta(h)}\sup_{v_0\in V_0}\inf_{v_h\in V_h}\left(\frac{\|v_0-v_h\|_V}{\|v_0\|_{V_0}}\right)\right\}=0,$$

to ensure the strong approximability. Indeed, the solution to the discrete mixed problem with $g = 0$ automatically belongs to the discrete kernel $\mathbb{K}_h$ and, as such, it can play the role of $v_{0,\mathbb{K}_h}$ in (4.5.15). Then, the following convergence result holds [51]:

Theorem 4.5.3. *Assume that* T *is a self-adjoint compact operator from* L *to* V *(cf. Assumption* (4.5.10)*). Assume moreover that the weak approximability of* Q_0 (4.5.14)*, the strong approximability of* V_0 (4.5.15) *and the coercivity on the discrete kernel (A2h) are fulfilled. Then, there exists* $r_3 : \mathbb{R}^+ \to \mathbb{R}^+$ *such that* $\lim_{h\to 0^+} r_3(h) = 0$ *and*

$$\sup_{f\in L} \frac{\|(\mathsf{T} - \mathsf{T}_h)f\|_V}{\|f\|_L} \leq r_3(h). \tag{4.5.16}$$

Proof. According to [51], the result is true with $r_3(h) = \left(1 + \frac{\|a\|}{\alpha'}\right) r_2(h) + \frac{1}{\alpha'} r_1(h)$. ◇

Once the convergence in operator norm is established, one can obtain error bounds on the eigenvalues and eigenspaces, thanks to techniques *à la Babuška-Osborn* (see [18, 48]). Let λ be an eigenvalue, and E_λ be the eigenspace associated to it. Let λ_h be the average of the discrete eigenvalues converging to λ, and E_{λ_h} the sum of the discrete eigenspaces associated to λ_h. One wants to measure the error $|\lambda - \lambda_h|$. As far as the eigenspaces are concerned, one wants to bound the *gap* between them. It can be defined mathematically as follows: given V_1 and V_2 two subspaces of V, introduce

$$\hat{\delta}(V_1, V_2) = \max(\delta(V_1, V_2), \delta(V_2, V_1)) \text{ where } \delta(V_1, V_2) = \sup_{\substack{v_1 \in V_1 \\ \|v_1\|_V = 1}} \inf_{v_2 \in V_2} \|v_1 - v_2\|_V.$$

We can prove the

Theorem 4.5.4. *There exists* $C > 0$ *such that*

$$|\lambda - \lambda_h| < C\,\varepsilon_\lambda(h)^2 \text{ and } \hat{\delta}(E_\lambda, E_{\lambda_h}) < C\varepsilon_\lambda(h), \tag{4.5.17}$$

with the approximation error

$$\varepsilon_\lambda(h) = \sup_{\substack{v \in E_\lambda \\ \|v\|_V = 1}} \inf_{v_h \in V_h} \|v - v_h\|_V. \tag{4.5.18}$$

4.6 T-coercivity

We refer to [60, 95] for details on the study of variational formulations in Hilbert spaces using T-coercivity. Consider first the variational problem, where V and W are two Hilbert spaces and $f \in W'$:

$$\text{Find } u \in V \text{ such that } \forall v \in W,\ a(u, v) = \langle f, v\rangle_W. \tag{4.6.1}$$

Classically, we know that Problem (4.6.1) is well-posed if $a(\cdot,\cdot)$ satisfies the stability and the solvability conditions of the so-called Banach–Nečas–Babuška (BNB) Theorem (see e.g. [148, thm 25.9]). For some models, one can also prove the well-posedness using the T-coercivity theory. Whereas the BNB Theorem relies on an abstract inf–sup condition, T-coercivity uses explicit inf–sup operators. We refer for instance to [61, 203, 57, 85, 58, 56] for problems with sign-changing coefficients, to [76, 73, 241] for integral equations, to [59] for interior transmission problems, to [176, 72, 92, 177, 171] for Helmholtz-like problems, to [A11, A6, 163] for the neutron diffusion equation, and to [A4] for the magnetostatic problem.

Definition 4.6.1 (`basic T-coercivity`)**.** *Let* V *and* W *be two Hilbert spaces and* $a(\cdot,\cdot)$ *be a continuous and bilinear form over* $V \times W$*. It is* T*-coercive if*

$$\exists T \in \mathcal{L}(V, W),\ \textit{bijective},\ \exists \alpha_T > 0,\ \forall v \in V,\ |a(v, Tv)| \geq \alpha_T \|v\|_V^2. \tag{4.6.2}$$

It is proved in [92, 85] that the T-coercivity condition is equivalent to the stability and solvability conditions of the BNB Theorem. Whereas the BNB Theorem relies on an abstract inf–sup condition, T-coercivity uses explicit inf–sup operators, both at the continuous and discrete levels.

Theorem 4.6.1 (`well-posedness [92, 85]`). *Let $a(\cdot,\cdot)$ be a continuous bilinear form on $V \times W$. The problem* (4.6.1) *is well-posed if, and only if, the form $a(\cdot,\cdot)$ is T-coercive in the sense of definition 4.6.1.*

Let us set $|||T||| := \sup_{v \in V\setminus\{0\}} \dfrac{\|T(v)\|_V}{\|v\|_V}$. Hence, we have: $a(u',v) \geq \dfrac{\alpha_T}{|||T|||}\|u'\|_V \, \|v\|_V$.
We recover the first Banach–Nečas–Babuška condition [148, Theorem 25.9, (BNB1)].
Thus, the basic T-coercivity gives an overestimate of the stability constant α given below:

$$\frac{\alpha_T}{|||T|||} \geq \alpha := \inf_{v \in V\setminus\{0\}} \sup_{u' \in V\setminus\{0\}} \frac{a(u',v)}{\|u'\|_V \, \|v\|_V}. \tag{4.6.3}$$

We can now use T-coercivity to build a continuous and coercive bilinear form:

Theorem 4.6.2 (`explicit T-coercivity`). *Let $a(\cdot,\cdot)$ be a continuous and bilinear form over $V \times W$. Suppose that the form $a(\cdot,\cdot)$ is T-coercive with $T \in \mathcal{L}(V,W)$. Then the bilinear form $a_T : V \times W$ such that for all $(u,v) \in V \times W$, $a_T(u,v) = a(u,T(v))$ is continuous and coercive. The variational formulation* (4.6.1) *is equivalent to:*

$$\textit{Find } u \in V \textit{ such that } \forall v \in V,\ a_T(u,v) = \langle f, T(v)\rangle_W. \tag{4.6.4}$$

Consider now a discretization of problem (4.6.1). We introduce $(V_h)_h$ and $(W_h)_h$ two sequences of finite-dimensional vector spaces, with $V_h \subset V$ and $W_h \subset W$ for all h (conforming approximation); we provide V_h with the norm $\|\cdot\|_V$, respectively $W_h \subset W$ with the norm $\|\cdot\|_W$. We suppose that $\dim V_h = \dim W_h$. By convention, the parameter h takes strictly positive values, and we have $\lim_{h\to 0}(\dim(V_h)) = +\infty$, with the approximability property recalled below.

Definition 4.6.2 (`approximability property, [60, Definition 3.6]`). *The family $(V_h)_h$ satisfies the approximability property if and only if*

$$\exists V^* \subset V \textit{ dense, } \forall h,\ \exists r_h : V^* \to V_h \textit{ such that } \forall v \in V^\star,\ \lim_{h\to 0} \|v - r_h v\|_V = 0.$$

The discretization of problem (4.6.1) reads:

$$\text{Find } u_h \in V_h \text{ such that } \forall v_h \in W_h,\ a_h(u_h,v_h) = \langle f_h, v_h\rangle_V, \tag{4.6.5}$$

where $a_h(\cdot,\cdot)$ and $f_h(\cdot)$ may differ from $a(\cdot,\cdot)_{|V_h\times W_h}$ and $f_h(\cdot)_{|W_h}$. We state a discrete T-coercivity condition.

Definition 4.6.3 (`discrete T-coercivity`). *The family of bilinear forms $(a_h)_h$ is uniformly T_h-coercive if and only if*

$$\exists \alpha_\dagger,\ \exists \beta_\dagger,\ \forall h > 0,\ \exists T_h \in \mathcal{L}(V_h,W_h), \quad |||T_h||| \leq \beta_\dagger \textit{ and } \forall v_h \in V_h,\ |a_h(v_h,T_h(v_h)| \geq \alpha_\dagger \|v_h\|^2_{V_h}.$$

We now establish the equivalence between the uniform discrete inf-sup condition, and the uniform T_h-coercivity of the forms $(a_h)_h$.

Theorem 4.6.3 (`discrete T-coercivity`). *Let be a family $(a_h)_h$ of bilinear forms, continuous and uniformly bounded. The three assertions below are equivalent:*

(i) The discrete problems (4.6.5) *are well-posed for any h.*

(ii) The family of bilinear forms $(a_h)_h$ satisfy a uniformly discrete inf-sup condition.

(iii) The family of bilinear forms $(a_h)_h$ is uniformly T_h-coercive.

The two key ideas initially developped in [85] are:

1. To discretize the variational formulation with (bijective) operator T cf. Problem (4.6.4).
2. To use the knowledge on operator T to derive the discrete operators $(T_h)_h$.

We can then use T-coercivity as an approximation tool. Given $h > 0$, let $N = \dim(V_h)$. Problem (4.6.5) is equivalent to:

$$\begin{cases} \text{Find } U \in \mathbb{C}^N \text{ s. t. } (\mathbb{A}U|W) = (F|W)\, \forall W \in \mathbb{C}^N, & \text{or:} \\ \text{Find } U \in \mathbb{C}^N \text{ s. t. } \mathbb{A}U = F. \end{cases}$$

Using $\mathbb{T}$ associated with T, Problem (4.6.5) is equivalent to:

$$\begin{cases} \text{Find } U \in \mathbb{C}^N \text{ s. t. } (\mathbb{A}U|\mathbb{T}V) = (F|\mathbb{T}V)\, \forall V \in \mathbb{C}^N, & \text{or:} \\ \text{Find } U \in \mathbb{C}^N \text{ s. t. } \mathbb{T}^*\mathbb{A}U = \mathbb{T}^*F. \end{cases}$$

According to the uniform T_h-coercivity assumption

$$\forall V \in \mathbb{C}^N, \quad |(\mathbb{T}^*\mathbb{A}V|V)| \geq \alpha_\dagger(\mathbb{M}V|V).$$

Samely, one can use $\mathbb{T}$ associated with $T_h(\cdot,\cdot)$ for the approximation of (4.6.4).

4.7 Discretization notations

The domain Ω is triangulated by a shape regular family of meshes $(\mathcal{T}_h)_h$, made up of L (closed) simplices K_ℓ, $\ell \in \mathcal{I}_K := \{1, \ldots, L\}$. Denoting by h_ℓ the diameter of K_ℓ, each mesh is indexed by the meshsize $h = \max_\ell h_\ell$. We denote by ρ_ℓ the diameter of the largest ball inscribed in K_ℓ, and we let: $\sigma_\ell = \frac{h_\ell}{\rho_\ell}$. When the $(\mathcal{T}_h)_h$ is a shape-regular triangulation sequence (see a.e. [147, def. 11.2]), it exists a constant $\sigma > 1$, called the shape regularity parameter, such that for all h, for all $\ell \in \mathcal{I}_K$, $\sigma_\ell \leq \sigma$. We denote by $\mathcal{I}_F$ denotes the index set of the faces, such that $\mathcal{F}_h := \bigcup_{f \in \mathcal{I}_F} F_f$ is the set of faces. Let $\mathcal{I}_F = \mathcal{I}_F^i \cup \mathcal{I}_F^b$, where $\forall f \in \mathcal{I}_F^i$, $F_f \in \overset{\circ}{\Omega}$ and $\forall f \in \mathcal{I}_F^b$, $F_f \in \partial\Omega$. For all $\ell \in \mathcal{I}_K$, we set $\mathcal{I}_{F,\ell}^{(i,b)} = \{f \in \mathcal{I}_F^{(i,b)} \,|\, F_f \in K_\ell\}$. We denote by $\mathcal{I}_S$ the index set of the vertices, such that $(S_j)_{j \in \mathcal{I}_S}$ is the set of vertices. Let $\mathcal{I}_S = \mathcal{I}_S^i \cup \mathcal{I}_S^b$, where $\forall j \in \mathcal{I}_S^i$, $S_j \in \overset{\circ}{\Omega}$ and $\forall j \in \mathcal{I}_S^b$, $S_j \in \partial\Omega$. Let introduce spaces of piecewise regular elements:
We set $\mathcal{P}_h H^1 = \left\{v \in L^2(\Omega)\,; \quad \forall \ell \in \mathcal{I}_K,\, v_{|K_\ell} \in H^1(K_\ell)\right\}$ and $\mathcal{P}_h\mathbf{H}^1 = [\mathcal{P}_h H^1]^d$.
We set $\mathcal{P}_h\mathbf{H}(\mathrm{div}) = \left\{\mathbf{v} \in \mathbf{L}^2(\Omega)\,; \quad \forall \ell \in \mathcal{I}_K,\, \mathbf{v}_{|K_\ell} \in \mathbf{H}(\mathrm{div}; K_\ell)\right\}$.
We define the element-wise operator ∇_h, $\boldsymbol{\nabla}_h$ and div_h.
Let $v \in \mathcal{P}_h H^1$. We define the jump of v through the facets. Let $f \in \mathcal{I}_F^i$ (resp. $f \in \mathcal{I}_F^b$) such that $F_f = \partial K_\ell \cap \partial K_{\ell'}$ (resp. $F_f \subset \partial K_\ell$) and $\mathbf{n}_f = \mathbf{n}_{f,\ell}$. The jump and the average are defined by:

$$[v]_f := v_{f,\ell} - v_{f,\ell'} \text{ (resp. } [v]_f = v_f := v_{f,\ell}) \text{ and } \{v\}_f = \frac{1}{2}\left(v_{f,\ell} + v_{f,\ell'}\right) \text{ (resp. } \{v\}_f = v_f := v_{f,\ell})$$

For all $D \subset \mathbb{R}^d$, $k \geq 0$, $P^k(D)$ is the set of order k polynomials on D, $\mathbf{P}^k(D) = (P^k(D))^d$, and:

$$\begin{array}{rrcl} \forall k \geq 0, & \mathcal{P}^k_{disc}(\mathcal{T}_h) & = & \left\{q \in L^2(\Omega);\quad \forall \ell \in \mathcal{I}_K,\, q_{|K_\ell} \in P^k(K_\ell)\right\}. \\ \forall k > 0, & \mathcal{P}^k(\mathcal{T}_h) & = & \mathcal{C}^0(\bar{\Omega}) \cap \mathcal{P}^k_{disc}(\mathcal{T}_h). \end{array}$$

We call π_0 the L^2-projection onto $\mathcal{P}^0_{disc}(\mathcal{T}_h)$: for all $q \in L^2(\Omega)$, $\pi_0(q) = \sum_{\ell \in \mathcal{I}_K} \underline{q}_\ell \mathbb{1}_{K_\ell}$. We use the notation $\mathcal{P}^0(\mathcal{T}_h)$ for $\mathcal{P}^0_{disc}(\mathcal{T}_h)$.

Chapter 5

Electromagnetism

The work presented in this Chapter concerns articles [A4, A1, A3], referenced below:

[A4] *Variational methods for solving numerically magnetostatic systems*,
P. Ciarlet Jr., E. Jamelot, Advances in Computational Mathematics **50**, 5, 2024.

[A1] *Solving electromagnetic eigenvalue problems in polyhedral domains with nodal finite elements*,
A. Buffa, P. Ciarlet Jr., E. Jamelot, Numerische Mathematik, **113**, 4, pp. 497–518, 2009.

[A3] *Continuous Galerkin methods for solving the time-dependent Maxwell equations in* $3D$ *geometries*,
P. Ciarlet Jr., E. Jamelot, Journal of Computational Physics **226**, 1, pp. 1122–1135, 2007.

This Chapter is organized as follow: we give details on Maxwell's equations and three different models in Section 5.1. Then in Section 5.2, we propose to solve the magnetostatic system using a mixed formulation obtained with the help of T-coercivity and discretized with edge finite elements [A4, §5]. To use continuous vector finite elements, we can rely on the so called *Weighted Regularization Method* (WRM) [116] that we recall in Section §5.3. Then, in Section 5.4, we study an augmented variational formulation obtained with the WRM [A4, §6]. We apply again the WRM, first to the electric eigenvalue problem [A1] in Section 5.5 and second to the propagation of the electric field in a waveguide [A3] in Section 5.6.

5.1 Maxwell's equations

In this Section, we refer to [16] for more precision. Let $\Omega \subset \mathbb{R}^3$ be the domain of study, that we will suppose to be a Lipschitz polyhedron. In the case of a cracked domain, the electromagnetic fields have a singular behavior near the crack tip. Discretization should be handled with care, as explained in [191] where the discretization of time-harmonic Maxwell equations is done using X-FEM for $2D$ edge elements. See also [220] for the analysis of eddy current.

The propagation of the electromagnetic fields in continuum media is described using the following vector real-valued functions defined in the space time $\mathbb{R} \times \Omega$: the electric field $\mathbf{E}$, the magnetic induction $\mathbf{B}$, the magnetic field $\mathbf{H}$ and the electric displacement $\mathbf{D}$, which are governed by Ampère's law, Faraday's law, Gauss' law and the absence of magnetic monopoles. They are two sources terms: the electrostatic charge density $\varrho \in \mathbb{R}$ and the current density $\mathbf{J} \in \mathbb{R}^3$. The differential Maxwell's equations read:

$$\partial_t \mathbf{D} - \mathbf{curl}\, \mathbf{H} = -\mathbf{J} \quad \text{(Ampère's law)}, \tag{5.1.1}$$

$$\partial_t \mathbf{B} + \mathbf{curl}\, \mathbf{E} = 0 \quad \text{(Faraday's law)}, \tag{5.1.2}$$

$$\operatorname{div} \mathbf{D} = \varrho \quad \text{(Gauss' law)}, \tag{5.1.3}$$

$$\operatorname{div} \mathbf{B} = 0 \quad \text{(absence of magnetic monopoles)}. \tag{5.1.4}$$

The differential charge conservation equation can be expressed as:

$$\operatorname{div} \mathbf{J} + \partial_t \varrho = 0. \tag{5.1.5}$$

Moreover, the behavior of the fields across an interface Σ is such that:

$$[\mathbf{D} \cdot \mathbf{n}]_\Sigma = \sigma_\Sigma, \quad [\mathbf{B} \cdot \mathbf{n}]_\Sigma = 0, \tag{5.1.6}$$

$$[\mathbf{n} \times \mathbf{E}]_\Sigma = 0, \quad [\mathbf{n} \times \mathbf{H}]_\Sigma = \mathbf{j}_\Sigma. \tag{5.1.7}$$

The data σ_Σ and $\mathbf{j}_\Sigma$ are the idealized surface charge density and the idealized surface current density such that $\sigma_\Sigma = \varrho \delta_\Sigma$. Let $\operatorname{div}_\Sigma$ be the surface divergence operator. Equation (5.1.5) yields:

$$\partial_t \sigma_\Sigma + \operatorname{div}_\Sigma \mathbf{j}_\Sigma + \left[\mathbf{J} \cdot \mathbf{n}_{|\Sigma}\right] = 0.$$

The system must be closed by adding constitutive relations relating $\mathbf{D}$ and $\mathbf{B}$ to $\mathbf{E}$ and $\mathbf{H}$. We suppose that the medium is perfect, that is, non-dispersive and anisotropic, or inhomogeneous, that is perfect and isotropic. In a perfect medium, the constitutive relations read as, for all $(t, \mathbf{x}) \in \mathbb{R} \times \Omega$:

$$\mathbf{D}(t, \mathbf{x}) = \underline{\varepsilon}(\mathbf{x}) \mathbf{E}(t, \mathbf{x}) \text{ and } \mathbf{B}(t, \mathbf{x}) = \underline{\mu}(\mathbf{x}) \mathbf{H}(t, \mathbf{x}),$$

where the constitutive parameters are the tensor functions $\underline{\varepsilon} \in \mathbb{R}^{3\times 3}$ and $\underline{\mu} \in \mathbb{R}^{3\times 3}$ which depend on the space variable $\mathbf{x}$. Equations (5.1.3)–(5.1.4) read:

$$\begin{aligned}
\underline{\varepsilon} \partial_t \mathbf{E} - \mathbf{curl}\, \underline{\mu}^{-1} \mathbf{B} &= -\mathbf{J} \quad \text{(Ampère's law)}, && (5.1.8)\\
\partial_t \mathbf{B} + \mathbf{curl}\, \mathbf{E} &= 0 \quad \text{(Faraday's law)}, && (5.1.9)\\
\operatorname{div} \underline{\varepsilon} \mathbf{E} &= \varrho \quad \text{(Gauss' law)}, && (5.1.10)\\
\operatorname{div} \mathbf{B} &= 0 \quad \text{(absence of magnetic monopoles)}. && (5.1.11)
\end{aligned}$$

For inhomogeneous medium, the constitutive parameters are scalar functions: $\varepsilon \in \mathbb{R}$ is called the electric permittivity and $\mu \in \mathbb{R}$ is called the magnetic permeability. When the medium is homogeneous, ε and μ are constant numbers. Last, in a conducting medium, the current density $\mathbf{J}$ and the electric field $\mathbf{E}$ are governed by Ohm's law: $\mathbf{J} = \underline{\sigma} \mathbf{E}$, where $\underline{\sigma} \in \mathbb{R}^3 \times \mathbb{R}^3$ is the tensor of conductivity which depends on the space variable $\mathbf{x}$.

Let us detail the three models we will consider.

5.1.1 Magnetostatic systems

The domain Ω can be simply connected or not [168]. This property is mathematically characterized by the value of the first Betti number, $\beta_1(\Omega)$: if $\beta_1(\Omega) = 0$, the domain Ω is simply connected and, if $\beta_1(\Omega) > 0$, it is non-simply connected. This means that we assume that one of the two conditions below holds:

- $\beta_1(\Omega) = 0$ *'for all vector fields* $\mathbf{v} \in \mathcal{C}^0(\Omega)$ *such that* $\mathbf{curl}\, \mathbf{v} = 0$ *in* Ω*, there exists* $p \in \mathcal{C}^1(\Omega)$ *such that* $\mathbf{v} = \nabla p$ *in* Ω*'*;
- $\beta_1(\Omega) > 0$ *'there exist* $I = \beta_1(\Omega)$ *non-intersecting, piecewise plane manifolds* $(\Sigma_j)_{j=1}^I$*, called cuts, with boundaries* $\partial\Sigma_i \subset \partial\Omega$*, such that, if one introduces the connected set* $\dot{\Omega} = \Omega \setminus \bigcup_{i=1}^I \Sigma_i$*, for all vector fields* $\mathbf{v} \in \mathcal{C}^0(\Omega)$ *such that* $\mathbf{curl}\, \mathbf{v} = 0$ *in* Ω*, there exists* $\dot{p} \in \mathcal{C}^1(\dot{\Omega})$ *such that* $\mathbf{v} = \nabla \dot{p}$ *in* $\dot{\Omega}$*'*.

Given $v \in L^2(\dot{\Omega})$ (resp. $\mathbf{v} \in \mathbf{L}^2(\dot{\Omega})$), we denote by $\widetilde{v}$ (resp. $\widetilde{\mathbf{v}}$) its continuation to $L^2(\Omega)$ (resp. $\mathbf{L}^2(\Omega)$).

The boundary $\partial\Omega$ is split into $K+1$ maximal connected components $(\Gamma_k)_{k=0}^K$, where $K = \beta_2(\Omega)$ is the second Betti number. We denote by Γ_0 the boundary of the unbounded component of $\mathbb{R}^3 \setminus \overline{\Omega}$.

Given $\xi \in \mathbb{L}^\infty(\Omega)$, we use the notation $\xi_+ = \|\xi\|_{\mathbb{L}^\infty(\Omega)}$. We say that $\xi \in \mathbb{L}^2(\Omega)$ is piecewise smooth if there exists a partition[1] $(\Omega_p^\xi)_{p=1}^P$ of Ω so that $\xi_{|\Omega_p^\xi} \in \mathbb{W}^{1,\infty}(\Omega_p^\xi)$ for $p \in \{1,\dots,P\}$. In this case, the partition is called compatible (with respect to ξ). Finally, for $\mathtt{s} > 0$, we introduce (cf. notations of §4.4.1):

$$\mathcal{P}H^{\mathtt{s}}(\Omega) = \{f \in L^2(\Omega) \text{ such that } f_{|\Omega_p^\xi} \in H^{\mathtt{s}}(\Omega_p^\xi),\ \forall p \in \{1,\dots,P\}\}.$$

Neglecting the time-variation of the electric displacement, one finds that the magnetic induction $\mathbf{B} \in \mathbf{L}^2(\Omega)$ is governed by the so-called *magnetostatic* system of equations. We refer to [63, chap. 6] or [35, chap. 3] for its definition. See also [40, 36, 236, 37, 38, 120] and References therein for related models.

$$\left\{\begin{array}{l} \text{Find } \mathbf{B} \in \mathbf{L}^2(\Omega) \text{ such that} \\ \qquad \mathbf{curl}(\mu^{-1}\mathbf{B}) = \mathbf{J} \text{ in } \Omega \\ \qquad \operatorname{div}\mathbf{B} = 0 \text{ in } \Omega \\ \qquad \mathbf{B}\cdot\mathbf{n} = 0 \text{ on } \partial\Omega \\ \langle \mathbf{B}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Sigma_i)} = c_i,\ \forall i \in \{1,\dots,I\} \end{array}\right. . \tag{5.1.12}$$

Above, the pair of data $(\mathbf{J}, (c_i)_{i=1}^I)$ is: $\mathbf{J} \in \mathbf{L}^2(\Omega)$ the electric current density in Ω; the numbers $(c_i)_{i=1}^I$ are related to current intensities flowing through some ad hoc exterior surfaces, see [35, §3.6]. It is also possible to use a formulation with the magnetic field $\mathbf{H} = \mu^{-1}\mathbf{B}$ as the main unknown. Because of the first equation, written $\mathbf{J} = \mathbf{curl}\,\mathbf{H}$ with $\mathbf{H} \in \mathbf{H}(\mathbf{curl}, \Omega)$, we note that $\mathbf{J} \in \mathbf{H}(\operatorname{div}; \Omega)$ with $\operatorname{div}\mathbf{J} = 0$ in Ω. Also, one may check that $\langle \mathbf{J}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_k)} = 0$, for all $k \in \{1,\dots,K\}$, as in Remark 3.4.2 in [16]. In what follows, we therefore consider any pair of data $(\mathbf{J}, (c_i)_{i=1}^I)$ with

$$\mathbf{J} \in \mathbf{H}(\operatorname{div}, \Omega) \text{ s.t. } \operatorname{div}\mathbf{J} = 0 \text{ in } \Omega,\ \langle \mathbf{J}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_k)} = 0,\ k \in \{1,\dots,K\}. \tag{5.1.13}$$

Mathematically, we assume that the permeability μ is a complex-valued, measurable, bounded, elliptic (see next Definition) and smooth or piecewise smooth tensor field. We observe that, in this case, the magnetic induction and the data are a priori complex-valued. We recall definitions and properties that can be found e.g. in [88, 87]. From now on, piecewise smoothness is understood with respect to a given compatible partition $(\Omega_p^\mu)_{p=1}^P$.

Definition 5.1.1. *The tensor-valued field $\xi \in \mathbb{L}^\infty(\Omega)$ is elliptic if*

$$\exists(\theta_\xi, \xi_-) \in \mathbb{R}\times\mathbb{R}_{>0},\quad \textit{a.e. in } \Omega,\ \forall \mathbf{z} \in \mathbb{C}^3,\quad \xi_-|\mathbf{z}|^2 \le \Re[e^{i\theta_\xi}\xi\mathbf{z}\cdot\overline{\mathbf{z}}]. \tag{5.1.14}$$

In (5.1.14), θ_ξ is called an ellipticity direction. We denote by Θ_ξ the set of admissible ellipticity directions

$$\Theta_\xi = \{\theta_\xi \in \mathbb{R},\ (\theta_\xi, \xi_-) \textit{ fulfills (5.1.14) for some } \xi_- \in \mathbb{R}_{>0}\}.$$

See [86, §3.1] for a discussion on this definition.

Proposition 5.1.1. *Let $\zeta \in \mathbb{L}^\infty(\Omega)$. If ζ is elliptic, one has $\zeta^{-1} \in \mathbb{L}^\infty(\Omega)$ with*

$$(\zeta^{-1})_+ \le \inf_{(\theta_\xi,\xi_-) \textit{ s.t. (5.1.14) with } \xi=\zeta \textit{ holds}} \big((\zeta_-)^{-1}\big).$$

Moreover, $\Theta_{\zeta^{-1}} = \{-\theta_\zeta,\ \theta_\zeta \in \Theta_\zeta\}$ and, given any $\theta_{\zeta^{-1}} \in \Theta_{\zeta^{-1}}$, one can choose $(\zeta^{-1})_- = \zeta_-(\zeta_+)^{-2}$ in (5.1.14) with $\xi = \zeta^{-1}$, where $(-\theta_{\zeta^{-1}}, \zeta_-)$ is such that (5.1.14) with $\xi = \zeta$ holds.

For conciseness, we assume that $0 \in \Theta_\mu$, so that $0 \in \Theta_{\mu^{-1}}$ according to the above (see footnote[2] page 38 for the case where $0 \notin \Theta_{\mu^{-1}}$). So, for the possibly nonhermitian elliptic tensor field μ, denoting by μ_-

[1] We recall that a partition of Ω is $(\Omega_p)_{p=1}^P$ such that

$$\Omega_p \text{ is a domain, for } p \in \{1,\dots,P\};\ \Omega_p \cap \Omega_q = \emptyset \text{ for } p \neq q;\ \overline{\Omega} = \cup_{p=1}^P \overline{\Omega_p}.$$

a value in (5.1.14) with $\xi = \mu$ for $\theta_\mu = 0$, one knows that $(\mu^{-1})_+ \leq (\mu_-)^{-1}$, and that $(\mu^{-1})_- = \mu_-(\mu_+)^{-2}$ is an admissible value in (5.1.14) with $\xi = \mu^{-1}$ for $\theta_{\mu^{-1}} = 0$. We use those values throughout the Chapter.

As a matter of fact, if the boundary is of class $\mathcal{C}^2$, using the theory developed e.g. in [150, 129], one ends up with convergence results that are identical to those we obtain.

By linearity, we see that solving the magnetostatic problem (5.1.12) can be done in two independent steps, by splitting the data: a first solve with the pair $(\mathbf{J}, 0)$ with a solution called $\mathbf{B}_J$, and a second one with the pair $(0, (c_i)_{i=1}^I)$, with a solution called $\mathbf{B}_c$. The total solution is then $\mathbf{B} = \mathbf{B}_J + \mathbf{B}_c$. This is this approach that we follow now. The pair of data is equal to $(\mathbf{J}, 0)$, with $\mathbf{J} \in \mathbf{L}^2(\Omega)$. In this case, we note that the magnetic induction fulfills the conditions

$$\mathbf{B}_J \in \mathbf{H}_0(\mathrm{div}, \Omega),\ \mathrm{div}\, \mathbf{B}_J = 0 \text{ in } \Omega \text{ and } \langle \mathbf{B}_J \cdot \mathbf{n}, 1 \rangle_{H^{1/2}(\Sigma_i)} = 0,\ \forall i \in \{1, \ldots, I\}.$$

According to e.g. [16, Theorem 3.5.1], there exists a vector potential $\mathbf{A} \in \mathbf{H}_0(\mathbf{curl}, \Omega)$ such that $\mathbf{B}_J = \mathbf{curl}\, \mathbf{A}$ in Ω. Moreover, one may characterize the vector potential by choosing a divergence-free, flux-free potential $\mathbf{A}$, i.e. $\mathrm{div}\, \mathbf{A} = 0$ in Ω and $\langle \mathbf{A} \cdot \mathbf{n}, 1 \rangle_{H^{1/2}(\Gamma_k)} = 0$, for all $k = 0, \ldots, K$. Obviously, the converse assertion is true: such a field $\mathbf{A}$ is such that $\mathbf{B}_J = \mathbf{curl}\, \mathbf{A}$ solves the above. So we conclude that, with the pair of data $(\mathbf{J}, 0)$, the magnetostatic system is equivalently recast as

$$\left\{ \begin{array}{rcl} \multicolumn{3}{c}{\text{Find } \mathbf{A} \in \mathbf{H}_0(\mathbf{curl}, \Omega) \text{ such that}} \\ \mathbf{curl}(\mu^{-1}\, \mathbf{curl}\, \mathbf{A}) & = & \mathbf{J} \text{ in } \Omega \\ \mathrm{div}\, \mathbf{A} & = & 0 \text{ in } \Omega \\ \langle \mathbf{A} \cdot \mathbf{n}, 1 \rangle_{H^{1/2}(\Gamma_k)} & = & 0,\ \forall k \in \{1, \ldots, K\} \end{array} \right. , \tag{5.1.15}$$

and then one defines $\mathbf{B}_J = \mathbf{curl}\, \mathbf{A}$.
We will study the resolution of this model with the help of T-coercivity and discretized with edge finite elements in §5.2; and with the Weighted Regularization Method in §5.4.

5.1.2 The electric eigenvalue problem

Consider the case where Ω is a resonator cavity, bounded by a perfect conductor. Hence: $\mathbf{J} = 0$, $\varrho = 0$, $\underline{\varepsilon} = \varepsilon_0$ and $\underline{\mu} = \mu_0$. The goal is to find eigenmodes of electromagnetic oscillations. Hence, we suppose that $\mathbf{E}(t, \mathbf{x}) = \Re(\mathbf{e}(\mathbf{x}) \exp(i\, \omega t))$ and $\mathbf{B}(t, \mathbf{x}) = \Re(\mathbf{b}(\mathbf{x}) \exp(i\, \omega t))$. We make the following abuse of notation: we write $\mathbf{E}$ and $\mathbf{B}$ instead for $\mathbf{e}$ and $\mathbf{b}$. Now $\mathbf{E}$ and $\mathbf{B}$ are complex-valued vectors. The electromagnetic eigenmodes $(\mathbf{E}, \mathbf{B}, \omega)$ are non-zero solutions to the time-harmonic Maxwell equations:

$$i\, \omega \mathbf{E} - c^2\, \mathbf{curl}\, \mathbf{B} = 0 \quad \text{(Ampère's law)}, \tag{5.1.16}$$

$$i\, \omega \mathbf{B} + \mathbf{curl}\, \mathbf{E} = 0 \quad \text{(Faraday's law)}, \tag{5.1.17}$$

together with Equations (5.1.10) and (5.1.11). The boundary conditions on $\partial\Omega$ are: $\mathbf{E} \times \mathbf{n} = 0$ and $\mathbf{B} \cdot \mathbf{n} = 0$. It is well known that one of the fields can be eliminated: multiplying (5.1.16) by $i\omega$ and adding $\mathbf{curl}$ of (5.1.17), we get a vector wavelike equation for $\mathbf{E}$, which reads $-\omega^2 \mathbf{E} + c^2\, \mathbf{curl}\, \mathbf{curl}\, \mathbf{E} = 0$. Similarly, there holds $-\omega^2 \mathbf{B} + c^2\, \mathbf{curl}\, \mathbf{curl}\, \mathbf{B} = 0$. Next, we we will study the electric eigenvalue problem below:

$$\left\{ \begin{array}{rcl} \multicolumn{3}{c}{\text{Find } \mathbf{E} \in \mathbf{H}_0(\mathbf{curl}, \Omega) \text{ and } \omega \text{ such that}} \\ c^2\, \mathbf{curl}\, \mathbf{curl}\, \mathbf{E} & = & \omega^2 \mathbf{E} \quad \text{in } \Omega \\ \mathrm{div}\, \mathbf{E} & = & 0 \quad \text{in } \Omega \end{array} \right. , \tag{5.1.18}$$

where we set: $\mathbf{H}_0(\mathbf{curl}, \Omega) := \{\mathbf{v} \in \mathbf{H}(\mathbf{curl}, \Omega) \,|\, \mathbf{v} \times \mathbf{n}_{|\partial\Omega} = 0\}$.
We will study the resolution of this model with the Weighted Regularization Method in §5.5.

5.1.3 Propagation in a waveguide

Consider the case where Ω is be a waveguide. In the vacuum, we set $\varepsilon = \varepsilon_0$, and $\mu = \mu_0$, which are such that $\varepsilon_0\,\mu_0\,c = 1$, where c is the light velocity. We suppose that the boundary $\partial\Omega$ is made of two disjoints parts $\partial\Omega = \overline{\Gamma}_A \cup \overline{\Gamma}_C$, where Γ_A is an artificial boundary and Γ_C is a perfectly conducting boundary, so that $\mathbf{E}\times\mathbf{n}_{|\Gamma_C} = 0$ and $\mathbf{B}\cdot\mathbf{n}_{|\Gamma_C} = 0$. We further split Γ_A into two parts Γ_A^i and Γ_A^a. On Γ_A^i, we model incoming plane waves, whereas we impose an absorbing boundary condition on Γ_A^a. Both can be modelled [17] as a Silver–Müller boundary condition on Γ_A: $(c\mathbf{B}+\mathbf{E}\times\mathbf{n})\times\mathbf{n} = c\mathbf{b}\times\mathbf{n}$, where $\mathbf{b}$ is given and equal to 0 in the case of an absorbing condition. If we differentiate with respect to time Equation (5.1.8) and add curl of quation (5.1.9)) to it, we get a vector wavelike equation for $\mathbf{E}$, which reads $\partial_{tt}^2\mathbf{E} + c^2\,\mathbf{curl}\,\mathbf{curl}\,\mathbf{E} = -\partial_t\mathbf{J}/\varepsilon_0$. Naturally, the initial value of $\partial_t\mathbf{E}$ is equal to $\mathbf{E}_1 := c^2\,\mathbf{curl}\,\mathbf{B}_0 - \mathbf{J}(0,\cdot)/\varepsilon_0$. Recall that $\mathbf{E}$ is also subject to a constraint on its divergence cf. (5.1.10), we have, using the trace of (5.1.9) on Γ_A, that $c^2\,\mathbf{curl}\,\mathbf{E}\times\mathbf{n} = c\,\partial_t\mathbf{E}\times\mathbf{n}\times\mathbf{n} - c^2\mathbf{b}\times\mathbf{n}$ on Γ_A. We thus consider the equivalent problem:
Find $\mathbf{E}$ *such that for all* $t\in]0,T[$*:*

$$\begin{aligned}
\partial_{tt}^2\mathbf{E} + c^2\,\mathbf{curl}\,\mathbf{curl}\,\mathbf{E} &= -\partial_t\mathbf{J}/\varepsilon_0 \quad \text{in } \Omega, \text{ for } t\in]0,T[, && (5.1.19)\\
\operatorname{div}\mathbf{E} &= \varrho/\varepsilon_0 \quad \text{in } \Omega, \text{ for } t\in]0,T[\,, && (5.1.20)\\
\mathbf{E}\times\mathbf{n}_{|\Gamma_C} &= 0, \text{ for } t\in]0,T[\,, && (5.1.21)\\
(c\,\mathbf{curl}\,\mathbf{E}\times\mathbf{n} - \partial_t\mathbf{E}\times\mathbf{n}\times\mathbf{n})_{|\Gamma_A} &= -c\,\partial_t\mathbf{b}\times\mathbf{n}_{|\Gamma_A} \quad \text{on } \Gamma_A, \text{ for } t\in]0,T[\,, && (5.1.22)
\end{aligned}$$

and satisfying the initial conditions:

$$\begin{aligned}
\mathbf{E}(0,\cdot) &= \mathbf{E}_0 \quad \text{in } \Omega, && (5.1.23)\\
\partial_t\mathbf{E}(0,\cdot) &= \mathbf{E}_1 \quad \text{in } \Omega. && (5.1.24)
\end{aligned}$$

Let us set:

$$\begin{aligned}
\mathbf{H}_{0,\Gamma_C}(\mathbf{curl},\Omega) &:= \{\mathbf{v}\in\mathbf{H}(\mathbf{curl},\Omega)\,|\,\mathbf{v}\times\mathbf{n}_{\partial\Omega}\in\mathbf{L}_t^2(\partial\Omega),\ \mathbf{v}\times\mathbf{n}_{|\Gamma_C}=0\},\ \text{where}\\
\mathbf{L}_t^2(\partial\Omega) &:= \{\mathbf{v}\in\mathbf{L}^2(\partial\Omega)\,|\,\mathbf{v}\cdot\mathbf{n}=0 \text{ a.e. }\};\\
\mathbf{X}_{N,\Gamma_C}(\Omega) &:= \mathbf{H}_{0,\Gamma_C}(\mathbf{curl},\Omega)\cap\mathbf{H}(\operatorname{div},\Omega);\\
\mathbf{H}(\operatorname{div}0,\Omega) &:= \{\mathbf{v}\in\mathbf{H}(\operatorname{div},\Omega)\,|\,\operatorname{div}\mathbf{v}=0\}.
\end{aligned} \qquad (5.1.25)$$

Notice that if $\Gamma_C = \partial\Omega$, $\mathbf{H}_{0,\Gamma_C}(\mathbf{curl},\Omega) = \mathbf{H}_0(\mathbf{curl},\Omega)$, and we use the notation $\mathbf{X}_0(\Omega) = \mathbf{X}_{N,\partial\Omega}(\Omega)$. Under suitable data assumption, it is known that $\mathbf{E}(t)\in\mathbf{X}_{N,\Gamma_C}(\Omega)$. We will study the resolution of this model with the Weighted Regularization Method in §5.6.

5.1.4 Edge or continuous finite elements?

In the three models above, the field belongs to $\mathbf{H}_0(\mathbf{curl},\Omega)$ or $\mathbf{H}_{0,\Gamma_C}(\mathbf{curl},\Omega)$. Thus, we can solve these problems using variational formulations set in $\mathbf{H}_{0,\Gamma_C}(\mathbf{curl},\Omega)$ discretized with classical edge finite elements [204, 199]. One does not have to know about the geometry of the domain (if it is convex or not) to implement the variational formulation.

On the other hand, it is advised, while solving the Maxwell-Vlasov system, to compute a continuous approximation of the field, especially when it is coupled to a particle-in-cell method [17, 45]. Moreover, continuous approximation provides spurious-free solutions and well-conditioned matrices [208]. Nevertheless, care must be taken when implementing the variational formulation if the domain is not convex.

To compute a continuous approximation, we consider fields in $\mathbf{H}(\mathbf{curl},\Omega)\cap\mathbf{H}(\operatorname{div},\Omega)$.

As it is difficult to build a conforming discretization in $\mathbf{H}(\operatorname{div}0,\Omega)\cap\mathbf{H}_{0,\Gamma_C}(\mathbf{curl},\Omega)$, the divergence-free condition on the field is usually prescribed as a natural condition. The field actually belongs to $\mathbf{X}_{N,\Gamma_C}(\Omega)$. According to [112], the inner norm and the semi-norm: $\|\mathbf{F}\|^2_{\mathbf{X}_0(\Omega)} := \|\,\mathbf{curl}\,\mathbf{F}\|^2_{\mathbf{L}^2(\Omega)} +$

$\|\operatorname{div}\mathbf{F}\|^2_{L^2(\Omega)}$ are equivalent norms on $\mathbf{X}_0(\Omega)$. Let us set $\mathbf{X}_0^R(\Omega) = \mathbf{X}_0(\Omega) \cap \mathbf{H}^1(\Omega)$. When Ω is convex, $\mathbf{X}_0^R(\Omega) = \mathbf{X}_0(\Omega)$, and we can use continuous vector finite elements. When Ω is not convex, $\mathbf{X}_0^R(\Omega)$ is a closed subspace of $\mathbf{X}_0(\Omega)$ [7]: $\mathbf{X}_0^R(\Omega)$ is not dense in $\mathbf{X}_0(\Omega)$: we must adapt the variational formulation to use continuous vector finite elements.

Stating that the domain Ω is not convex amounts to considering that $\partial\Omega$ includes a set of reentrant edges (resp. corners in $2D$) $\mathcal{E}$, with dihedral angles $(\pi/\alpha_e)_{e\in\mathcal{E}}$ such that $\alpha_e \in]\frac{1}{2}, 1[$. Let $\mathbf{X}_0^S(\Omega) \subset \mathbf{X}^0(\Omega)$ be the singular part of $\mathbf{X}_0(\Omega)$, i.e. $\mathbf{X}_0^S(\Omega) \not\subset \mathbf{H}^1(\Omega)$. In $2D$, this space is of finite dimension equal to the number of reentrant corners, $\operatorname{card}(\mathcal{E})$, and its basis is known. One can compute precisely $\mathbf{X}_0(\Omega)$-fields with *the Singular Complement Method*, which consists in splitting the field into an explicit singular part plus a regular part, computed with continuous finite elements, cf. [62, 173, 157, B2, A8].

In $3D$, the space $\mathbf{X}_0^S(\Omega)$ is not of finite dimension. If the domain is prismatic or axisymmetric, one can still compute singular fields with continuous finite elements, using the Fourier Singular Complement Method, see [102] for axisymmetric domains.

There exist at least two other approaches, as described in [91]. First one can choose to "relax" the boundary conditions, which are then treated as *natural boundary conditions*. This allows to overcome the density problem mentioned above, by solving Maxwell equations in larger functional spaces e.g. $\mathbf{X}(\Omega) := \{\mathbf{v} \in \mathbf{H}(\mathbf{curl}, \Omega) \cap \mathbf{H}(\operatorname{div}, \Omega) \,|\, \mathbf{v} \times \mathbf{n}_{|\partial\Omega} \in \mathbf{L}^2_t(\partial\Omega)\}$, in which the subspace of $\mathbf{H}^1$-smooth fields is dense [99, 113]. The boundary condition will be taken into account in an *augmented variational formulation* [91, B2].

Second, one can control the divergence of the field in weighted Sobolev spaces. Costabel and Dauge proposed in [116] to take the divergence explicitly into account in the measure of the vector potentials, see [91, A8, A3, A1, 101, 103, 207] for applications in electromagnetism in a homogeneous material. This is the first method that was designed to approximate successfully singular fields in $3D$ domains by continuous finite elements. In most of the above cited References, the problem is solved in a simplified geometrical setting, and, the assumptions on the permeability μ are more restrictive than those we set in §5.1.1. For other methods using the same mathematical framework, we refer to [65, 138, 54, 20, 137].

Concerning the triangulation, the polyhedron $\overline{\Omega}$ is triangulated by a shape regular family of meshes $(\mathcal{T}_h)_h$, cf. Section 4.7. For heterogeneous materials, we assume that the meshes are conforming with the compatible partition $(\Omega_p^\mu)_{p=1}^P$ and/or with respect to the cuts. In other words, for all h, for all $\ell \in \mathcal{I}_K$, there exists $p \in \{1, \ldots, P\}$ such that $K_\ell \subset \overline{\Omega_p}$ and $int(K_\ell) \cap (\cup_{i=1}^I \Sigma_i) = \emptyset$. This is a realistic assumption that allows to simplify the numerical analysis presented here.

5.2 Magnetostatic systems, explicit T-coercivity [A4, §5]

Given data $\mathbf{J}$ that fulfill assumption (5.1.13), we now build a first variational formulation, that is equivalent to (5.1.15). Let us start by simple observations. We notice that, given $\mathbf{u} \in \mathbf{H}_0(\mathbf{curl}, \Omega)$, $\mathbf{u}$ is such that $\mathbf{curl}(\mu^{-1}\,\mathbf{curl}\,\mathbf{u}) = \mathbf{J}$ in Ω if, and only if,

$$\forall \mathbf{v} \in \mathbf{H}_0(\mathbf{curl}, \Omega), \qquad (\mu^{-1}\,\mathbf{curl}\,\mathbf{u}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} = (\mathbf{J}, \mathbf{v})_{\mathbf{L}^2(\Omega)}. \tag{5.2.1}$$

Then, let $H^1_{\partial\Omega}(\Omega) = \big\{q \in H^1(\Omega) \text{ such that } q_{|\Gamma_0} = 0,\ q_{|\Gamma_k} \in \mathbb{C},\ \forall k \in \{1, \ldots, K\}\big\}$. One uses the norm $q \mapsto \|\nabla q\|_{\mathbf{L}^2(\Omega)}$ in $H^1_{\partial\Omega}(\Omega)$. One checks easily that, given $\mathbf{u} \in \mathbf{L}^2(\Omega)$, $\mathbf{u}$ is such that $\operatorname{div}\mathbf{u} = 0$ in Ω and $\langle \mathbf{u}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_k)} = 0$, for $k \in \{1, \ldots, K\}$ if, and only if,

$$\forall q \in H^1_{\partial\Omega}(\Omega), \qquad (\mathbf{u}, \nabla q)_{\mathbf{L}^2(\Omega)} = 0. \tag{5.2.2}$$

Hence, it is equivalent to solve (5.1.15), or the variational formulation:

$$\left\{\begin{array}{l} \text{Find } \mathbf{A} \in \mathbf{H}_0(\mathbf{curl},\Omega) \text{ such that} \\ (\mu^{-1}\,\mathbf{curl}\,\mathbf{A},\mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} = (\mathbf{J},\mathbf{v})_{\mathbf{L}^2(\Omega)} \quad \forall \mathbf{v} \in \mathbf{H}_0(\mathbf{curl},\Omega) \\ \qquad\qquad (\mathbf{A},\nabla q)_{\mathbf{L}^2(\Omega)} = 0 \qquad\qquad \forall q \in H^1_{\partial\Omega}(\Omega) \end{array}\right. . \tag{5.2.3}$$

5.2.1 A symmetrized formulation with artificial pressure

Classically, we build a second variational formulation, that is symmetrized via the introduction of an artificial pressure. As a matter of fact, given $\gamma > 0$, we aim at solving the symmetrized, complex-valued saddle-point formulation:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{A},p) \in \mathbf{H}_0(\mathbf{curl},\Omega) \times H^1_{\partial\Omega}(\Omega) \text{ such that} \\ (\mu^{-1}\,\mathbf{curl}\,\mathbf{A},\mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} + \gamma\,(\mathbf{v},\nabla p)_{\mathbf{L}^2(\Omega)} = (\mathbf{J},\mathbf{v})_{\mathbf{L}^2(\Omega)} \quad \forall \mathbf{v} \in \mathbf{H}_0(\mathbf{curl},\Omega) \\ \qquad\qquad \gamma\,(\mathbf{A},\nabla q)_{\mathbf{L}^2(\Omega)} = 0 \qquad\qquad \forall q \in H^1_{\partial\Omega}(\Omega) \end{array}\right. . \tag{5.2.4}$$

Theorem 5.2.1 (`[A4, Theorem 2]`). *Let $\gamma > 0$ be given, and let the data $\mathbf{J}$ fulfill assumption (5.1.13). Then, the vector potential $\mathbf{A}$ solves (5.1.15) if, and only if, $(\mathbf{A},0)$ solves (5.2.4).*

Proof. If $\mathbf{A}$ solves (5.2.3), then $(\mathbf{A},0)$ solves (5.2.4). Thanks to the assumption (5.1.13) on $\mathbf{J}$, any solution $(\mathbf{A},p)$ to (5.2.4) is such that $p = 0$. We note that, since $p \in H^1_{\partial\Omega}(\Omega)$ it holds that $\mathbf{v} = \nabla p \in \mathbf{H}_0(\mathbf{curl},\Omega)$. Inserting in (5.2.4), we have $\gamma\,\|\nabla p\|^2_{\mathbf{L}^2(\Omega)} = (\mathbf{J},\nabla p)_{\mathbf{L}^2(\Omega)} = -(\operatorname{div}\mathbf{J},p)_{\mathbf{L}^2(\Omega)} + \sum_{k=0}^{K} \langle \mathbf{A}\cdot\mathbf{n},p\rangle_{H^{1/2}(\Gamma_k)} = 0.$ ⋄

5.2.2 Using T-coercivity

Let us apply the theory of T-coercivity to solve the variational formulation (5.2.4), cf. [26]. Let $\mathcal{V} = \mathbf{H}_0(\mathrm{curl},\Omega) \times H^1_{\partial\Omega}(\Omega)$, endowed with the norm

$$\|(\mathbf{v},q)\|_{\mathcal{V}} = \left(\|\mathbf{v}\|^2_{\mathbf{H}(\mathbf{curl},\Omega)} + \|\nabla q\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}}.$$

On can rewrite (5.2.4) as: Find $(\mathbf{A},p) \in \mathcal{V}$ such that for all $(\mathbf{v},q) \in \mathcal{V}$

$$a((\mathbf{A},p),(\mathbf{v},q)) = \ell((\mathbf{v},q)), \tag{5.2.5}$$

where $a(\cdot,\cdot)$ (resp. $\ell(\cdot)$) is the following continuous sesquilinear form on $\mathcal{V}\times\mathcal{V}$, (resp. anti-linear continuous form on $\mathcal{V}$)

$$\begin{array}{rcl} a((\mathbf{u},p),(\mathbf{v},q)) &=& (\mu^{-1}\,\mathbf{curl}\,\mathbf{u},\mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} + \gamma\,(\mathbf{u},\nabla p)_{\mathbf{L}^2(\Omega)} + \gamma\,(\mathbf{u},\nabla q)_{\mathbf{L}^2(\Omega)} \\ \ell((\mathbf{v},q)) &=& (\mathbf{J},\mathbf{v})_{\mathbf{L}^2(\Omega)}. \end{array}$$

We shall prove that the form $a(\cdot,\cdot)$ is T-coercive. In $\mathbf{H}_0(\mathbf{curl},\Omega)$, we introduce the orthogonal subspace to the range of the gradient operator from $H^1_{\partial\Omega}(\Omega)$:

$$\mathbf{H}_0(\mathbf{curl},\Omega) = \nabla[H^1_{\partial\Omega}(\Omega)] \overset{\perp}{\oplus} \mathbf{K}^-_N(\Omega), \tag{5.2.6}$$

where the orthogonality is understood with respect to the "natural" inner product $(\cdot,\cdot)_{\mathbf{H}(\mathbf{curl},\Omega)}$. The decomposition (5.2.6) is usually called a Helmholtz decomposition. It is easily checked that

$$\begin{array}{rcl} \mathbf{K}^-_N(\Omega) &=& \left\{\mathbf{k} \in \mathbf{H}_0(\mathbf{curl},\Omega) \text{ s.t. } \forall q \in H^1_{\partial\Omega}(\Omega),\ (\mathbf{k},\nabla q)_{\mathbf{L}^2(\Omega)} = 0\right\}, \\ &=& \left\{\mathbf{k} \in \mathbf{H}_0(\mathbf{curl},\Omega) \text{ s.t. } \operatorname{div}\mathbf{k} = 0 \text{ in } \Omega,\ \langle \mathbf{k}\cdot\mathbf{n},1\rangle_{H^{1/2}(\Gamma_k)} = 0,\ \forall k\right\}. \end{array}$$

Above, the first line is an instance of the famous double orthogonality property in electromagnetism. In our case, it holds with respect to the "natural" $\mathbf{H}(\mathbf{curl},\Omega)$ and $\mathbf{L}^2(\Omega)$ inner products. This double orthogonality property is crucial to establish T-coercivity. Before we proceed, we recall a first Weber inequality [244]. There exists $C_W > 0$ such that, for all $\mathbf{v} \in \mathbf{H}_0(\mathbf{curl},\Omega) \cap \mathbf{H}(\mathrm{div},\Omega)$, one has

$$\|\mathbf{v}\|_{\mathbf{L}^2(\Omega)} \leq C_W \left\{ \|\,\mathbf{curl}\,\mathbf{v}\|_{\mathbf{L}^2(\Omega)} + \|\,\mathrm{div}\,\mathbf{v}\|_{L^2(\Omega)} + \sum_{k=1}^{K} |\langle \mathbf{v}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_k)}| \right\}.$$

It follows that, for all $\mathbf{k} \in \mathbf{K}_N^-(\Omega)$, one has the bound

$$\|\mathbf{k}\|_{\mathbf{L}^2(\Omega)} \leq C_W \,\|\,\mathbf{curl}\,\mathbf{k}\|_{\mathbf{L}^2(\Omega)}. \tag{5.2.7}$$

Theorem 5.2.2 (`[A4, Theorem 3]`). *The form $a(\cdot,\cdot)$ is T-coercive.*

Proof. We decompose $\mathbf{u}$ into $\mathbf{u} = \mathbf{k} + \nabla\phi_u$, with $\mathbf{k}_u \in \mathbf{K}_N^-(\Omega)$ and $\phi_u \in H^1_{\partial\Omega}(\Omega)$, and we choose $\mathbf{v}^\star = \mathbf{k}_u + \nabla p$, respectively $q^\star = \phi_u$. In particular: $\mathbf{curl}\,\mathbf{v}^\star = \mathbf{curl}\,\mathbf{k}_u = \mathbf{curl}\,\mathbf{u}$. We get, using the double orthogonality properties,

$$a((\mathbf{u},p),(\mathbf{v}^\star,q^\star)) = (\mu^{-1}\,\mathbf{curl}\,\mathbf{k}_u, \mathbf{curl}\,\mathbf{k}_u)_{\mathbf{L}^2(\Omega)} + \gamma\,\|\nabla p\|^2_{\mathbf{L}^2(\Omega)} + \gamma\,\|\nabla\phi_u\|^2_{\mathbf{L}^2(\Omega)}. \tag{5.2.8}$$

By assumption, $0 \in \Theta_{\mu^{-1}}$, with $(\mu^{-1})_- = \mu_-(\mu_+)^{-2}$.[2] So one finds that

$$\begin{aligned} \Re\left(a((\mathbf{u},p),(\mathbf{v}^\star,q^\star))\right) &\geq \alpha^\star \left\{ \|\mathbf{k}_u\|^2_{\mathbf{H}(\mathbf{curl},\Omega)} + \|\nabla p\|^2_{\mathbf{L}^2(\Omega)} + \|\nabla\phi_u\|^2_{\mathbf{L}^2(\Omega)} \right\} \\ &= \alpha^\star \left\{ \|\mathbf{k}_u + \nabla\phi_u\|^2_{\mathbf{H}(\mathbf{curl},\Omega)} + \|\nabla p\|^2_{\mathbf{L}^2(\Omega)} \right\} = \alpha^\star\,\|(\mathbf{u},p)\|^2_{\mathcal{V}} \end{aligned}$$

where $\alpha^\star = \min\left((\mu^{-1})_-(1+C_W^2)^{-1}, \gamma\right) = \min\left(\mu_-(\mu_+)^{-2}(1+C_W^2)^{-1}, \gamma\right) > 0$.
Let $(\mathbf{u},p) \in \mathcal{V}$. Using the above notation, we note that $\mathbf{v}^\star = \mathbf{k}_u + \nabla p$ is the orthogonal decomposition (5.2.6) of $\mathbf{v}^\star \in \mathbf{H}_0(\mathbf{curl},\Omega)$ since $\mathbf{k}_u \in \mathbf{K}_N^-(\Omega)$ and $q^\star \in H^1_{\partial\Omega}(\Omega)$, so that $T(\mathbf{v}^\star,q^\star) = (\mathbf{k}_u + \nabla\phi_u, p) = (\mathbf{u},p)$. Hence, $T\circ T = I_{\mathcal{V}}$, and T is a bijective operator. This concludes the proof. ◇

According to the abstract T-coercivity theory, one has the

Corollary 5.2.1. *The variational formulation (5.2.5) is well-posed.*

5.2.3 A perturbed approach

Let T be defined as in the proof of theorem 5.2.2: $T(\mathbf{v},q) = (\mathbf{k}_v + \nabla q, \phi_v)$. Since T is bijective, the variational formulation (5.2.5) is equivalent to:

$$\text{Find } (\mathbf{A},p) \in \mathcal{V} \text{ such that for all } (\mathbf{v},q) \in \mathcal{V}, \quad a((\mathbf{A},p),T(\mathbf{v},q)) = \ell(T(\mathbf{v},q)). \tag{5.2.9}$$

By design, the form $a' : ((\mathbf{u},p),(\mathbf{v},q)) \mapsto a((\mathbf{u},p),T(\mathbf{v},q))$ is coercive. Using $T(\mathbf{v},q) = (\mathbf{k}_v + \nabla q, \phi_v)$ one can compute its expression:

$$a'((\mathbf{u},p),(\mathbf{v},q)) = (\mu^{-1}\,\mathbf{curl}\,\mathbf{u}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} + \gamma\,(\nabla q, \nabla p)_{\mathbf{L}^2(\Omega)} + \gamma\,(\nabla\phi_u, \nabla\phi_v)_{\mathbf{L}^2(\Omega)}.$$

Computationally speaking, the first two terms are explicit, while one has to know the gradient parts of $\mathbf{u}$ and $\mathbf{v}$ to compute the last term.
Then, we note that, according to (5.1.13), the data $\mathbf{J}$ is orthogonal to $\nabla[H^1_{\partial\Omega}(\Omega)]$ in $\mathbf{L}^2(\Omega)$. So, the

[2] If $0 \notin \Theta_{\mu^{-1}}$, one chooses some $\theta \in \Theta_{\mu^{-1}}$ to solve (5.2.5) with the tilted forms

$$\begin{aligned} &a((\mathbf{u},p),(\mathbf{v},q)) = \exp(i\theta)\,(\mu^{-1}\,\mathbf{curl}\,\mathbf{u}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} + \gamma\,(\mathbf{v},\nabla p)_{\mathbf{L}^2(\Omega)} + \gamma\,(\mathbf{u},\nabla q)_{\mathbf{L}^2(\Omega)} \\ &\ell((\mathbf{v},q)) = \exp(i\theta)\,(\mathbf{J},\mathbf{v})_{\mathbf{L}^2(\Omega)}, \end{aligned}$$

and the proof can be carried out similarly. Furthermore, one can check easily that, if $0 \notin \Theta_{\mu^{-1}}$, all proofs given hereafter still hold by considering appropriately tilted forms.

expression of $\ell' : (\mathbf{v}, q) \mapsto \ell(T(\mathbf{v}, q))$ is

$$\ell'((\mathbf{v}, q)) = (\mathbf{J}, \mathbf{k}_v + \nabla q)_{\mathbf{L}^2(\Omega)} = (\mathbf{J}, \mathbf{k}_v + \nabla q + \nabla(\phi_v - q))_{\mathbf{L}^2(\Omega)} = (\mathbf{J}, \mathbf{v})_{\mathbf{L}^2(\Omega)} = \ell((\mathbf{v}, q)).$$

Hence one may rewrite (5.2.9) as the variational formulation:

$$\text{Find } (\mathbf{A}, p) \in \mathcal{V} \text{ such that for all } (\mathbf{v}, q) \in \mathcal{V}, \quad a'((\mathbf{A}, p), (\mathbf{v}, q)) = \ell((\mathbf{v}, q)).$$

Letting $\mathbf{v} = 0$ above, we note that $p = 0$, and we end up with the new variational formulation:

$$\text{Find } \mathbf{A} \in \mathbf{H}_0(\mathbf{curl}, \Omega) \text{ such that for all } \mathbf{v} \in \mathbf{H}_0(\mathbf{curl}, \Omega), \quad c(\mathbf{A}, \mathbf{v}) = (\mathbf{J}, \mathbf{v})_{\mathbf{L}^2(\Omega)}, \tag{5.2.10}$$

with $c(\mathbf{u}, \mathbf{v}) = (\mu^{-1} \, \mathbf{curl}\, \mathbf{u}, \mathbf{curl}\, \mathbf{v})_{\mathbf{L}^2(\Omega)} + \gamma \, (\nabla \phi_u, \nabla \phi_v)_{\mathbf{L}^2(\Omega)}$.

In the original variational formulation (5.2.5), all terms are explicitly computable, but the form $a(\cdot, \cdot)$ is not coercive. Moreover, the solution is made up of two parts, a physical part $\mathbf{A}$ and and artificial part p. On the other hand, in the new variational formulation (5.2.10), the second term in the expression of $c(\cdot, \cdot)$ term is not explicitly known, while the form $c(\cdot, \cdot)$ is coercive, and only the physical part $\mathbf{A}$ of the solution remains. At the discrete level, forgetting for a moment the non-explicit term in $c(\cdot, \cdot)$, the cost of linear solvers should be higher in the discrete version of (5.2.5) than in the discrete version of (5.2.10). Below, we propose a perturbed approach, that allows one to recover explicit terms in the definition of the coercive sesquilinear form, at the expense of solving an inexact problem with a "small" perturbation.

We observe that the solution $\mathbf{A}$ is *independent* of the value of the parameter γ. So, a natural idea is to choose a "small" value of γ, and to add a perturbation in the order of γ. How so? Let us introduce $c_\gamma(\mathbf{u}, \mathbf{v}) = c(\mathbf{u}, \mathbf{v}) + \gamma(\mathbf{k}_u, \mathbf{k}_v)_{\mathbf{L}^2(\Omega)}$. Using again the orthogonality properties, we find the following expression to the perturbed sesquilinear form

$$c_\gamma(\mathbf{u}, \mathbf{v}) = (\mu^{-1} \, \mathbf{curl}\, \mathbf{u}, \mathbf{curl}\, \mathbf{v})_{\mathbf{L}^2(\Omega)} + \gamma \, (\mathbf{u}, \mathbf{v})_{\mathbf{L}^2(\Omega)}.$$

Remark 5.2.1. *Since by assumption $0 \in \Theta_{\mu^{-1}}$, the form c_γ is coercive for all $\gamma > 0$. We note that, for γ smaller than $(\mu^{-1})_-$, the largest coercivity constant is equal to γ, while the norm remains bounded by $(\mu^{-1})_+$.*

Then, we are solving the perturbed variational formulation

$$\text{Find } \mathbf{A}_\gamma \in \mathbf{H}_0(\mathbf{curl}, \Omega) \text{ such that for all } \mathbf{v} \in \mathbf{H}_0(\mathbf{curl}, \Omega), \quad c_\gamma(\mathbf{A}_\gamma, \mathbf{v}) = (\mathbf{J}, \mathbf{v})_{\mathbf{L}^2(\Omega)}. \tag{5.2.11}$$

In general $\mathbf{A}_\gamma \neq \mathbf{A}$ since

$$\mathbf{curl}(\mu^{-1} \, \mathbf{curl}\, \mathbf{A}_\gamma) + \gamma \mathbf{A}_\gamma = \mathbf{J} \text{ in } \Omega, \tag{5.2.12}$$

This perturbed variational formulation corresponds to a model that has been considered in [216, 139, 104], in a simplified geometrical setting, that is in a simply connected domain, with a connected boundary. Also, the assumptions on the permeability μ are more restrictive in those references, namely it is a real-valued symmetric tensor field. The next proposition is proved in [216, 139] in the "simpler" configuration, this is the reason why we provide a sketched proof.

Proposition 5.2.1 (`[A4, Proposition 5]`). *For all $\gamma > 0$, $\mathbf{A}_\gamma \in \mathbf{K}_N^-(\Omega)$.*
Furthermore, letting $C' = (C_W)^3 \, (\mu_+)^4 (\mu_-)^{-2}$, one has the estimates

$$\| \mathbf{curl}(\mathbf{A}_\gamma - \mathbf{A}) \|_{\mathbf{L}^2(\Omega)} \leq C' \gamma \, \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)} \text{ and } \|\mathbf{A}_\gamma - \mathbf{A}\|_{\mathbf{L}^2(\Omega)} \leq C' C_W \gamma \, \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}. \tag{5.2.13}$$

Proof. Given $q \in H^1_{\partial\Omega}(\Omega)$, we choose the test-field $\mathbf{v} = \nabla q$ in (5.2.11). By assumption on the data, cf. (5.1.13), we find $(\mathbf{A}_\gamma, \nabla q)_{\mathbf{L}^2(\Omega)} = 0$. Hence, $\mathbf{A}_\gamma \in \mathbf{K}_N^-(\Omega)$. In addition, according to the Weber inequality (5.2.7), it follows that $\|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \leq$

$C_W \| \mathbf{curl}\, \mathbf{A}_\gamma \|_{\mathbf{L}^2(\Omega)}$. Thanks to $0 \in \Theta_{\mu^{-1}}$, then using the definition of c_γ and finally taking $\mathbf{v} = \mathbf{A}_\gamma$ in (5.2.11), we find that

$$\begin{aligned} \mu_-(\mu_+)^{-2} \| \mathbf{curl}\, \mathbf{A}_\gamma \|^2_{\mathbf{L}^2(\Omega)} &\leq \Re\left((\mu^{-1}\, \mathbf{curl}\, \mathbf{A}_\gamma, \mathbf{curl}\, \mathbf{A}_\gamma)_{\mathbf{L}^2(\Omega)} \right) \leq \Re\left(c_\gamma(\mathbf{A}_\gamma, \mathbf{A}_\gamma) \right) \leq \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)} \|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \\ &\leq C_W \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)} \| \mathbf{curl}\, \mathbf{A}_\gamma \|_{\mathbf{L}^2(\Omega)}. \end{aligned}$$

It follows that $\| \mathbf{curl}\, \mathbf{A}_\gamma \|_{\mathbf{L}^2(\Omega)} \leq C_W (\mu_+)^2 (\mu_-)^{-1} \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}$, and $\|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \leq (C_W)^2 (\mu_+)^2 (\mu_-)^{-1} \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}$.
Comparing (5.2.3a) to (5.2.11), we find that $\forall \mathbf{v} \in \mathbf{H}_0(\mathbf{curl}, \Omega)$, $(\mu^{-1}\, \mathbf{curl}(\mathbf{A}_\gamma - \mathbf{A}), \mathbf{curl}\, \mathbf{v})_{\mathbf{L}^2(\Omega)} + \gamma\, (\mathbf{A}_\gamma, \mathbf{v})_{\mathbf{L}^2(\Omega)} = 0$. Using $\mathbf{v} = (\mathbf{A}_\gamma - \mathbf{A}) \in \mathbf{K}^-_N(\Omega)$ above, we now find

$$\begin{aligned} \mu_-(\mu_+)^{-2} \| \mathbf{curl}(\mathbf{A}_\gamma - \mathbf{A}) \|^2_{\mathbf{L}^2(\Omega)} &\leq \Re\left((\mu^{-1}\, \mathbf{curl}(\mathbf{A}_\gamma - \mathbf{A}), \mathbf{curl}(\mathbf{A}_\gamma - \mathbf{A}))_{\mathbf{L}^2(\Omega)} \right), \\ &\leq \gamma\, \|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \|\mathbf{A}_\gamma - \mathbf{A}\|_{\mathbf{L}^2(\Omega)}, \\ &\leq C_W\, \gamma\, \|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \| \mathbf{curl}(\mathbf{A}_\gamma - \mathbf{A}) \|_{\mathbf{L}^2(\Omega)}, \end{aligned}$$

so that $\| \mathbf{curl}(\mathbf{A}_\gamma - \mathbf{A}) \|_{\mathbf{L}^2(\Omega)} \leq C_W\, (\mu_+)^2 (\mu_-)^{-1}\, \gamma\, \|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \leq (C_W)^3\, (\mu_+)^4 (\mu_-)^{-2}\, \gamma\, \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}$, i.e. the first inequality in (5.2.13). According to (5.2.7), we conclude that $\|\mathbf{A}_\gamma - \mathbf{A}\|_{\mathbf{L}^2(\Omega)} \leq (C_W)^4\, (\mu_+)^4 (\mu_-)^{-2}\, \gamma\, \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}$, which is the second inequality in (5.2.13). ⋄

5.2.4 Approximation by edge finite elements

Below, we propose to discretize the perturbed problem (5.2.11) using edge finite elements. The discrete solution will prove to be a suitable approximation of the vector potential $\mathbf{A}$, under appropriate conditions on the perturbation parameter γ. Recall that we are looking for an approximation of $\mathbf{B}_J = \mathbf{curl}\, \mathbf{A}$, so we focus on finding a suitable approximation, and suitable approximation results, for $\mathbf{curl}\, \mathbf{A}$. In this respect, compared to [216, 139, 104], the process is simplified, since one is looking for approximation results for both $\mathbf{A}$ and $\mathbf{curl}\, \mathbf{A}$ in those Refs.

We choose the Nédélec's first family of edge finite elements [204, 199] to define finite dimensional subspaces $(\mathbf{V}_h)_h$ of $\mathbf{H}_0(\mathbf{curl}, \Omega)$. Below, we consider first-order finite elements. In principle, the results can be extended to higher-order finite elements without any difficulty. So, let $\mathcal{R}_1(K_\ell)$ be the vector space of polynomials on K_ℓ defined by $\mathcal{R}_1(K_\ell) = \{\mathbf{v} \in \mathbf{P}^1(K_\ell)\ :\ \mathbf{v}(\mathbf{x}) = \mathbf{a} + \mathbf{b} \times \mathbf{x},\ \mathbf{a}, \mathbf{b} \in \mathbb{C}^3\}$. For given h, one introduces $\mathbf{V}_h = \{\mathbf{v}_h \in \mathbf{H}_0(\mathbf{curl}, \Omega)\ :\ \mathbf{v}_{h|K_\ell} \in \mathcal{R}_1(K_\ell),\ \forall \ell \in \mathcal{I}_K\}$. We introduce Π^{curl}_h, the classical global Nédélec interpolant in $\mathbf{H}_0(\mathbf{curl}, \Omega)$ with values in $\mathbf{V}_h$ [204, 199]. According to Lemma 5.1 of [39], we have the

Proposition 5.2.2. *Assume that* $\mathbf{v} \in \mathcal{P}\mathbf{H}^{\mathtt{s}}(\Omega)$ *and* $\mathbf{curl}\, \mathbf{v} \in \mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)$ *for some* $\mathtt{s} > 1/2$, $\mathtt{s}' > 0$. *Then one can define* $\Pi^{curl}_h \mathbf{v}$ *and, in addition, one has the interpolation result:*

$$\|\mathbf{v} - \Pi^{curl}_h \mathbf{v}\|_{\mathbf{H}(\mathbf{curl}, \Omega)} \lesssim h^{\min(\mathtt{s}, \mathtt{s}', 1)} \{ \|\mathbf{v}\|_{\mathcal{P}\mathbf{H}^{\mathtt{s}}(\Omega)} + \| \mathbf{curl}\, \mathbf{v} \|_{\mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)} \}. \tag{5.2.14}$$

Since we are interested in the approximation of the curl of $\mathbf{A}$, we now refine (5.2.14). We introduce Π^{div}_h the classical global Raviart-Thomas interpolation operator in $\mathbf{H}_0(\mathrm{div}, \Omega)$ with values in $\mathbf{W}_h$ [213, 204], where $(\mathbf{W}_h)_h$ are designed with the help of $\mathbf{H}(\mathrm{div}, \Omega)$-conforming, first-order finite element spaces. Here, $\mathbf{W}_h = \{\mathbf{w}_h \in \mathbf{H}_0(\mathrm{div}, \Omega)\ :\ \mathbf{w}_{h|K_\ell} \in \mathcal{D}_1(K_\ell),\ \forall \ell \in \mathcal{I}_K\}$, with $\mathcal{D}_1(K_\ell)$ the vector space of polynomials on K defined by $\mathcal{D}_1(K_\ell) = \{\mathbf{v} \in \mathbf{P}_1(K_\ell)\ :\ \mathbf{v}(\mathbf{x}) = \mathbf{a} + b\mathbf{x},\ \mathbf{a} \in \mathbb{C}^3,\ b \in \mathbb{C}\}$. Using the results of Chapter 5 in [199], we recall that $\forall \mathbf{v} \in \mathbf{H}_0(\mathbf{curl}, \Omega)$ s.t. $\Pi^{curl}_h \mathbf{v}$ exists, $\Pi^{div}_h(\mathbf{curl}\, \mathbf{v}) = \mathbf{curl}(\Pi^{curl}_h \mathbf{v})$. Then, applying Lemma 3.3 of [34] (interpolation error for Raviart-Thomas discretization) together with the previous property, we have

Proposition 5.2.3. *Assume that* $\mathbf{v} \in \mathbf{H}_0(\mathbf{curl}, \Omega)$ *and* $\mathbf{curl}\, \mathbf{v} \in \mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)$ *for some* $\mathtt{s}' > 0$. *Then if one can define* $\Pi^{curl}_h \mathbf{v}$, *one has the interpolation result:*

$$\| \mathbf{curl}(\mathbf{v} - \Pi^{curl}_h \mathbf{v}) \|_{\mathbf{L}^2(\Omega)} \lesssim h^{\min(\mathtt{s}', 1)} \| \mathbf{curl}\, \mathbf{v} \|_{\mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)}. \tag{5.2.15}$$

Given a perturbation parameter $\gamma > 0$ and a meshsize h, the discrete variational formulation of the

perturbed problem (5.2.11) is then

$$\text{Find } \mathbf{A}_\gamma^h \in \mathbf{V}_h \text{ such that for all } \mathbf{v}_h \in \mathbf{V}_h, \quad c_\gamma(\mathbf{A}_\gamma^h, \mathbf{v}_h) = (\mathbf{J}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}. \tag{5.2.16}$$

To carry out the numerical analysis, we rely on a variant of the Céa lemma.

Proposition 5.2.4 (`[A4, Proposition 8]`). *For all $\gamma > 0$ and all h, one has the estimate*

$$\begin{aligned} &\|\operatorname{\mathbf{curl}}(\mathbf{A}_\gamma - \mathbf{A}_\gamma^h)\|_{\mathbf{L}^2(\Omega)} \\ &\quad \le \inf_{\mathbf{v}_h \in \mathbf{V}_h} \left\{ \gamma^{1/2} \frac{\mu_+}{\sqrt{2}(\mu_-)^{1/2}} \|\mathbf{A}_\gamma - \mathbf{v}_h\|_{\mathbf{L}^2(\Omega)} + \frac{\mu_+^2}{\mu_-^2} \|\operatorname{\mathbf{curl}}(\mathbf{A}_\gamma - \mathbf{v}_h)\|_{\mathbf{L}^2(\Omega)} \right\}. \end{aligned} \tag{5.2.17}$$

Remark 5.2.2. *If the permittivity μ is a hermitian tensor field, then c_γ defines an inner product. Denoting $\|\cdot\|_\gamma : \mathbf{v} \mapsto (\gamma\|\mathbf{v}\|^2_{\mathbf{L}^2(\Omega)} + (\mu^{-1}\operatorname{\mathbf{curl}}\mathbf{v}, \operatorname{\mathbf{curl}}\mathbf{v})_{\mathbf{L}^2(\Omega)})^{1/2}$ the associated norm, one has the "straightforward" estimate for all $\gamma > 0$ and all h,*

$$\|\mathbf{A}_\gamma - \mathbf{A}_\gamma^h\|_\gamma \le \inf_{\mathbf{v}_h \in \mathbf{V}_h} \|\mathbf{A}_\gamma - \mathbf{v}_h\|_\gamma, \tag{5.2.18}$$

and it holds that $\|\operatorname{\mathbf{curl}}(\mathbf{A}_\gamma - \mathbf{A}_\gamma^h)\|_{\mathbf{L}^2(\Omega)} \le \mu_+ \inf_{\mathbf{v}_h \in \mathbf{V}_h} \|\mathbf{A}_\gamma - \mathbf{v}_h\|_\gamma$.

Using (5.2.13) and (5.2.17), we find

$$\|\operatorname{\mathbf{curl}}(\mathbf{A} - \mathbf{A}_\gamma^h)\|_{\mathbf{L}^2(\Omega)} \lesssim \inf_{\mathbf{v}_h \in \mathbf{V}_h} \left\{ \gamma^{1/2} \|\mathbf{A}_\gamma - \mathbf{v}_h\|_{\mathbf{L}^2(\Omega)} + \|\operatorname{\mathbf{curl}}(\mathbf{A}_\gamma - \mathbf{v}_h)\|_{\mathbf{L}^2(\Omega)} \right\} + \gamma \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}. \tag{5.2.19}$$

To bound the infimum, we would like to use $\mathbf{v}_h = \Pi_h^{curl} \mathbf{A}_\gamma$. This is possible if $\mathbf{A}_\gamma$ is sufficiently smooth, in the sense of Proposition 5.2.2.

Let us apply the results of §4.4.2 to our solution $\mathbf{A}_\gamma$.

Proposition 5.2.5. *Let $\mathbf{A}_\gamma$ be the solution to (5.2.11). Let $\mathtt{s} = 1$ if $\sigma_{Dir} = 1$, and $\mathtt{s} \in]1/2, \sigma_{Dir}[$ else. Then, $\mathbf{A}_\gamma \in \mathbf{H}^{\mathtt{s}}(\Omega)$, and $\|\mathbf{A}_\gamma\|_{\mathbf{H}^{\mathtt{s}}(\Omega)} \lesssim_{\mathtt{s}} \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}$.*

Proof. Since $\mathbf{A}_\gamma \in \mathbf{K}_N^-(\Omega) \subset \mathbf{H}_0(\mathbf{curl}, \Omega)$, with the help of the regular-gradient splitting (4.4.3), we find:

$$\mathbf{A}_\gamma = \mathbf{A}_{reg} + \nabla q_A \text{ in } \Omega, \text{ with } \mathbf{A}_{reg} \in \mathbf{H}^1(\Omega),\ q_A \in H_0^1(\Omega), \text{ and } \|\mathbf{A}_{reg}\|_{\mathbf{H}^1(\Omega)} + \|q_A\|_{H^1(\Omega)} \lesssim \|\mathbf{A}_\gamma\|_{\mathbf{H}(\mathbf{curl},\Omega)}.$$

By construction, $\Delta q_A = -\operatorname{div}\mathbf{A}_{reg}$ in Ω, with $\|\operatorname{div}\mathbf{A}_{reg}\|_{L^2(\Omega)} \le \sqrt{3}\|\mathbf{A}_{reg}\|_{\mathbf{H}^1(\Omega)}$. According to the shift theorem, using the definition of the limit regularity exponent σ_{Dir}, we find that $\nabla q_A \in \mathbf{H}^{\mathtt{s}}(\Omega)$, with $\|\nabla q_A\|_{\mathbf{H}^{\mathtt{s}}(\Omega)} \lesssim_{\mathtt{s}} \|\mathbf{A}_{reg}\|_{\mathbf{H}^1(\Omega)}$. Hence, $\mathbf{A}_\gamma \in \mathbf{H}^{\mathtt{s}}(\Omega)$, with the estimate $\|\mathbf{A}_\gamma\|_{\mathbf{H}^{\mathtt{s}}(\Omega)} \lesssim_{\mathtt{s}} \|\mathbf{A}_\gamma\|_{\mathbf{H}(\mathbf{curl},\Omega)}$, and the first Weber inequality yields $\|\mathbf{A}_\gamma\|_{\mathbf{H}^{\mathtt{s}}(\Omega)} \lesssim_{\mathtt{s}} \|\operatorname{\mathbf{curl}}\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)}$. Finally, we recall that $\|\operatorname{\mathbf{curl}}\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \lesssim \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}$ (proof of proposition 5.2.1), so the conclusion follows. ⋄

There holds a second Weber inequality (cf. [88, Theorem 2.11]), namely there exists $C'_W > 0$ such that, for all $\mathbf{v} \in \mathbf{H}(\mathbf{curl}, \Omega) \cap \mathbf{H}_0(\operatorname{div}\mu, \Omega)$, one has

$$\|\mathbf{v}\|_{\mathbf{L}^2(\Omega)} \le C'_W \left\{ \|\operatorname{\mathbf{curl}}\mathbf{v}\|_{\mathbf{L}^2(\Omega)} + \|\operatorname{div}\mu\mathbf{v}\|_{L^2(\Omega)} + \sum_{i=1}^{I} |\langle \mu\mathbf{v}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Sigma_i)}| \right\}.$$

One has a second shift theorem, which we state directly for elements of $\mathbf{H}(\mathbf{curl}, \Omega) \cap \mathbf{H}_0(\operatorname{div}\mu, \Omega)$, cf. [119, 94]. It is derived thanks to another regular-gradient splitting, which is omitted.
Let $\mathbf{v} \in \mathbf{H}(\mathbf{curl}, \Omega) \cap \mathbf{H}_0(\operatorname{div}\mu, \Omega)$. There exists $\sigma_{Neu}(\mu) \in]0, 1]$ such that

$$\mathbf{H}(\mathbf{curl}, \Omega) \cap \mathbf{H}_0(\operatorname{div}\mu, \Omega) \subset \mathcal{P}\dot{\mathbf{H}}^{\sigma_{Neu}(\mu)}(\Omega) = \bigcap_{0 \le \mathtt{s}' < \sigma_{Neu}(\mu)} \mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega), \tag{5.2.20}$$

with continuous embedding.

Applying the two results together, one concludes that, for every $\mathtt{s}' \in]0, \sigma_{Neu}(\mu)[$, one has

$$\|\mathbf{v}\|_{\mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)} \lesssim_{\mathtt{s}'} \left\{ \|\,\mathbf{curl}\,\mathbf{v}\|_{\mathbf{L}^2(\Omega)} + \|\,\mathrm{div}\,\mu\mathbf{v}\|_{L^2(\Omega)} + \sum_{i=1,I} |\langle \mu\mathbf{v}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Sigma_i)}| \right\}.$$

Regarding the curl of $\mathbf{A}_\gamma$, we find the results below.

Proposition 5.2.6. *Let $\mathbf{A}_\gamma$ be the solution to (5.2.11). Let $\mathtt{s}' = 1$ if $\sigma_{Neu}(\mu) = 1$, and $\mathtt{s}' \in]0, \sigma_{Neu}(\mu)[$ else. Then, $\mathbf{curl}\,\mathbf{A}_\gamma \in \mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)$, and $\|\,\mathbf{curl}\,\mathbf{A}_\gamma\|_{\mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)} \lesssim_{\mathtt{s}'} (1+\gamma)\,\|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}$.*

Proof. The field $\mathbf{H}_\gamma = \mu^{-1}\,\mathbf{curl}\,\mathbf{A}_\gamma$ is such that $\mathbf{H}_\gamma \in \mathbf{H}(\mathbf{curl}, \Omega)$ with $\mathbf{curl}\,\mathbf{H}_\gamma = \mathbf{J} - \gamma\mathbf{A}_\gamma$ in Ω according to (5.2.12). Also, $\mathbf{H}_\gamma \in \mathbf{H}_0(\mathrm{div}\,\mu, \Omega)$, with $\mathrm{div}(\mu\mathbf{H}_\gamma) = 0$ in Ω. Using the second shift theorem for $\mathbf{H}_\gamma$, we find $\mathbf{H}_\gamma \in \mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)$. Next, because $\mu\mathbf{H}_\gamma \in \mathbf{curl}\,[\mathbf{H}_0(\mathbf{curl}, \Omega)]$, one has always $\langle \mu\mathbf{H}_\gamma\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Sigma_i)} = 0$ for $i = 1, I$ (cf. remark 3.5.2. in [16]). So we have $\|\mathbf{H}_\gamma\|_{\mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)} \lesssim_{\mathtt{s}'} \|\,\mathbf{curl}\,\mathbf{H}_\gamma\|_{\mathbf{L}^2(\Omega)}$. Going back to $\mathbf{curl}\,\mathbf{H}_\gamma = \mathbf{J} - \gamma\mathbf{A}_\gamma$ in Ω, we infer

$$\|\mathbf{H}_\gamma\|_{\mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)} \lesssim_{\mathtt{s}'} \left\{ \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)} + \gamma\|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \right\}.$$

Recall that the partition $(\Omega_p^\mu)_{p=1,P}$ is chosen so that $\mu_{|\Omega_p^\mu} \in (W^{1,\infty}(\Omega_p^\mu))^{3\times 3}$ for $p = 1, P$. Because multiplying by $\mu_{|\Omega_p^\mu}$ is stable in $\mathbf{H}^{\mathtt{s}'}(\Omega_p^\mu)$ for $p = 1, P$ [55], we find that $\mathbf{curl}\,\mathbf{A}_\gamma = \mu\mathbf{H}_\gamma \in \mathcal{P}\mathbf{H}^{\mathtt{s}'}(\Omega)$, with

$$\|\,\mathbf{curl}\,\mathbf{A}_\gamma\|_{\mathbf{PH}^{\mathtt{s}'}(\Omega)} \lesssim_{\mathtt{s}'} \left\{ \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)} + \gamma\|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \right\}.$$

To conclude, we recall that $\|\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} \lesssim \|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}$ (proof of proposition 5.2.1). ◇

As announced, one may use $\mathbf{v}_h = \Pi_h^{curl}\mathbf{A}_\gamma$ in (5.2.19).

Theorem 5.2.3. *For $\gamma > 0$ and $h > 0$, let $\mathbf{A}_\gamma^h$ be the solution to (5.2.16).*
Then, for $\mathtt{s} = 1$ if $\sigma_{Dir} = 1$, and $\mathtt{s} \in]1/2, \sigma_{Dir}[$ else, and for $\mathtt{s}' = 1$ if $\sigma_{Neu}(\mu) = 1$, and $\mathtt{s}' \in]0, \sigma_{Neu}(\mu)[$ else, one has the error estimate

$$\|\,\mathbf{curl}(\mathbf{A} - \mathbf{A}_\gamma^h)\|_{\mathbf{L}^2(\Omega)} \lesssim_{\mathtt{s},\mathtt{s}'} \left(\gamma + \gamma^{1/2}(1+\gamma)h^{\min(\mathtt{s},\mathtt{s}')} + (1+\gamma)h^{\mathtt{s}'}\right)\|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}. \tag{5.2.21}$$

Proof. Let us bound the infimum in (5.2.19). Combining the interpolation errors (5.2.14) and (5.2.15) with the last two propositions, we find

$$\begin{aligned}\inf_{\mathbf{v}_h\in\mathbf{V}_h} &\{\gamma^{1/2}\|\mathbf{A}_\gamma - \mathbf{v}_h\|_{\mathbf{L}^2(\Omega)} + \|\,\mathbf{curl}(\mathbf{A}_\gamma - \mathbf{v}_h)\|_{\mathbf{L}^2(\Omega)}\}\\ &\lesssim \gamma^{1/2}\|\mathbf{A}_\gamma - \Pi_h^{curl}\mathbf{A}_\gamma\|_{\mathbf{L}^2(\Omega)} + \|\,\mathbf{curl}(\mathbf{A}_\gamma - \Pi_h^{curl}\mathbf{A}_\gamma)\|_{\mathbf{L}^2(\Omega)},\\ &\lesssim_{\mathtt{s},\mathtt{s}'} \gamma^{1/2}(1+\gamma)h^{\min(\mathtt{s},\mathtt{s}')}\|\mathbf{J}\|_{\mathbf{L}^2(\Omega)} + (1+\gamma)h^{\mathtt{s}'}\|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}.\end{aligned}$$

Using (5.2.19) leads to the claim. ◇

To conclude this part, we now consider that γ is a function of the meshsize h (with strictly positive values), that is $\gamma = \gamma(h)$, and we solve

$$\text{Find } \mathbf{A}_h \in \mathbf{V}_h \text{ such that for all } \mathbf{v}_h \in \mathbf{V}_h, \quad c_{\gamma(h)}(\mathbf{A}_h, \mathbf{v}_h) = (\mathbf{J}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}. \tag{5.2.22}$$

Corollary 5.2.2. *Assume that $h \in]0, 1]$. Let $\gamma = \gamma(h)$ be such that $\gamma(h) \lesssim h^{\sigma_{Neu}(\mu)}$.*
For $\mathtt{s}' = 1$ if $\sigma_{Neu}(\mu) = 1$, and $\mathtt{s}' \in]0, \sigma_{Neu}(\mu)[$ else, one has the error estimate

$$\|\,\mathbf{curl}(\mathbf{A} - \mathbf{A}_h)\|_{\mathbf{L}^2(\Omega)} \lesssim_{\mathtt{s}'} h^{\mathtt{s}'}\,\|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}. \tag{5.2.23}$$

Proof. Since in theorem 5.2.3 one has $\mathtt{s} > 1/2$, it holds $h^{\min(\mathtt{s},\mathtt{s}')} \le h^{\min(1/2,\mathtt{s}')}$. Hence from (5.2.21) one gets the bound:

$$\|\,\mathbf{curl}(\mathbf{A} - \mathbf{A}_\gamma^h)\|_{\mathbf{L}^2(\Omega)} \lesssim_{\mathtt{s}'} \left(\gamma + \gamma^{1/2}(1+\gamma)h^{\min(1/2,\mathtt{s}')} + (1+\gamma)h^{\mathtt{s}'}\right)\|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}.$$

Let us find the dominant term between the parentheses. If $\mathtt{s}' < 1/2$,

$$\gamma(h) + \gamma(h)^{1/2}(1+\gamma(h))h^{\min(1/2,\mathtt{s}')} \lesssim h^{\sigma_{Neu}(\mu)} + h^{\sigma_{Neu}(\mu)/2+\mathtt{s}'} = O(h^{\mathtt{s}'}).$$

If $\mathtt{s}' \geq 1/2$, taking into account that $\sigma_{Neu}(\mu)/2 + 1/2 > \sigma_{Neu}(\mu)$, one has

$$\gamma(h) + \gamma(h)^{1/2}(1+\gamma(h))h^{\min(1/2,\mathtt{s}')} \lesssim h^{\sigma_{Neu}(\mu)} + h^{\sigma_{Neu}(\mu)/2+1/2} = O(h^{\sigma_{Neu}(\mu)}).$$

Since $\gamma(h)h^{\mathtt{s}'} = o(h^{\mathtt{s}'})$, the dominant term is $h^{\mathtt{s}'}$. $\diamond$

Hence, choosing $\gamma(h) \lesssim h^{\sigma_{Neu}(\mu)}$ and going back to the magnetic field $\mathbf{B}_J$, it is guaranteed that

$$\|\mathbf{B}_J - \mathbf{curl}\,\mathbf{A}_h\|_{\mathbf{H}(\mathrm{div},\Omega)} \lesssim_{\mathtt{s}'} h^{\mathtt{s}'}\,\|\mathbf{J}\|_{\mathbf{L}^2(\Omega)}.$$

According to the definition of $\mathtt{s}'$, one concludes that the a priori error is equal to $O(h)$ if $\sigma_{Neu}(\mu) = 1$, and almost like $O(h^{\sigma_{Neu}(\mu)})$ if $\sigma_{Neu}(\mu) < 1$, since the result is $O(h^{\mathtt{s}'})$ for all $\mathtt{s}' < \sigma_{Neu}(\mu)$ in the latter case.

Because the form $c_{\gamma(h)}$ is coercive, the linear system equivalent to (5.2.22) is expected to be simpler to solve than the one resulting from the discretization of the saddle-point variational formulation (5.2.4) or (5.2.5). Morever, according to remark 5.2.1, it is advised to take the "largest" possible values of $\gamma(h)$ to improve the conditioning of the matrix. In other words, one should take $\gamma(h) \eqsim h^{\sigma_{Neu}(\mu)}$, see corollary 5.2.2.

One can find a number of numerical experiments in [216, 139, 104] to support the claims, and to show the robustness of the approach. We propose below a numerical illustration with a manufactured solution.

5.3 The Weighted Regularization Method

This approach aims to design a mathematical framework that can be efficiently approximated numerically by continuous vector finite elements. When Ω is non-convex[3], following [116], the divergence of the fields is evaluated in a weighted L^2 space. Recall that $\mathcal{E}$ is the non-empty set of reentrant edges of $\partial\Omega$. Let d be the distance to $\mathcal{E}$: $d(x) = dist(x, \cup_{e\in\mathcal{E}}\bar{e})$. Consider w_α a smooth non-negative function of x, that depends on a real parameter α. The (simplified) weight w_α is chosen to behave locally as d^α in the neighborhood of reentrant edges and corners, and is bounded above and below by a strictly positive constant outside a fixed neighborhood of $\mathcal{E}$. Let:

$$L^2_\alpha(\Omega) = \{q \in L^2_{\mathrm{loc}}(\Omega) \text{ such that } w_\alpha\, q \in L^2(\Omega)\},$$

endowed with the natural norm $\|\cdot\|_{L^2_\alpha(\Omega)} : q \mapsto \|w_\alpha\, q\|_{L^2(\Omega)}$, and associated scalar product $(\cdot,\cdot)_{L^2_\alpha(\Omega)}$. Then, we introduce a subspace of $\mathbf{H}_0(\mathbf{curl},\Omega)$, endowed with the natural norm:

$$\mathbf{X}^\alpha_N(\Omega) = \{\mathbf{v} \in \mathbf{H}_0(\mathbf{curl},\Omega) \text{ s.t. } \mathrm{div}\,\mathbf{v} \in L^2_\alpha(\Omega)\}, \quad \|\mathbf{v}\|_{\mathbf{X}^\alpha_N(\Omega)} = \Big(\|\mathbf{v}\|^2_{\mathbf{H}(\mathbf{curl},\Omega)} + \|\mathrm{div}\,\mathbf{v}\|^2_{L^2_\alpha(\Omega)}\Big)^{1/2}.$$

In the absence of weight, we write $\mathbf{X}_N(\Omega) := \mathbf{X}^0_N(\Omega)$.

Introducing the value $\alpha_{\min} = 1 - \sigma_{Dir} \in]0, 1/2[$, one has that for all $\alpha \in]\alpha_{\min}, 1[$, $\mathbf{X}^\alpha_N(\Omega) \cap \mathbf{H}^1(\Omega)$ is dense in $\mathbf{X}^\alpha_N(\Omega)$ according to [116, Theorem 5.1]. This result is paramount to obtain the basic approximability property for a conforming finite element method in $\mathbf{X}^\alpha_N(\Omega)$. So, we assume from now on that $\alpha \in]\alpha_{\min}, 1[$.

Among the noticeable properties proven in [116], we observe that $H^1_0(\Omega) \subset L^2_{-\alpha}(\Omega)$ (with dense embedding), see bottom of page 249 of *loco citato*. For $k \in \{1,\ldots,K\}$, let q_k be characterized by

$$\text{Find } q_k \in H^1_{\partial\Omega}(\Omega) \text{ such that } -\Delta q_k = 0 \text{ in } \Omega, q_{k|\Gamma_m} = \delta_{km} \text{ for } m \in \{1,\ldots,K\}. \tag{5.3.1}$$

[3]Non-convexity always occurs when the boundary $\partial\Omega$ is not connected, that is $K \geq 1$.

By construction, one has the orthogonal decomposition

$$H^1_{\partial\Omega}(\Omega) = H^1_0(\Omega) \overset{\perp}{\oplus} \mathrm{span}((q_\ell)_{m=1}^K). \tag{5.3.2}$$

One can check that 1 belongs to $L^2_{-\alpha}(\Omega)$ so, for $k \in \{1,\dots,K\}$, we also have that $q_k \in L^2_{-\alpha}(\Omega)$. As a consequence, the fluxes $(\langle \mathbf{v}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_\ell)})_{m=1}^K$ are meaningful for all $\mathbf{v} \in \mathbf{X}^\alpha_N(\Omega)$, and equal to $(\langle \mathbf{v}\cdot\mathbf{n}, q_m\rangle_{H^{1/2}(\partial\Omega)})_{m=1}^K$.
Then, following for instance [16, proof of Theorem 3.4.3] and using the embedding of $L^2_\alpha(\Omega)$ in $H^{-1}(\Omega)$, one can prove that

$$\|\cdot\|_{\mathbf{X}^\alpha_N(\Omega)} : \mathbf{v} \mapsto \left(\|\mathbf{curl}\,\mathbf{v}\|^2_{\mathbf{L}^2(\Omega)} + \|\mathrm{div}\,\mathbf{v}\|^2_{L^2_\alpha(\Omega)} + \sum_{m=1}^K |\langle \mathbf{v}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)}|^2 \right)^{1/2}$$

defines a norm, that is equivalent to the natural norm, in $\mathbf{X}^\alpha_N(\Omega)$. We let $(\cdot,\cdot)_{\mathbf{X}^\alpha_N(\Omega)}$ be the associated scalar product.

5.4 Magnetostatic systems, continuous finite elements [A4, §6]

5.4.1 Augmented variational formulation

Going back to the magnetic system (5.1.15), it is obvious that its divergence-free solution $\mathbf{A}$ automatically belongs to $\mathbf{X}^\alpha_N(\Omega)$. Let $\mathbf{v} \in \mathbf{X}^\alpha_N(\Omega)$. Then, adding the contributions

$$\begin{aligned} &(\mu^{-1}\,\mathbf{curl}\,\mathbf{A}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} = (\mathbf{curl}\,\mu^{-1}\,\mathbf{curl}\,\mathbf{A}, \mathbf{v})_{\mathbf{L}^2(\Omega)} = (\mathbf{J},\mathbf{v})_{\mathbf{L}^2(\Omega)}\,; \\ &(\mathrm{div}\,\mathbf{A}, \mathrm{div}\,\mathbf{v})_{L^2_\alpha(\Omega)} = 0\,; \quad \sum_{m=1}^K \langle \mathbf{A}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)} \overline{\langle \mathbf{v}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)}} = 0\,; \end{aligned}$$

we find that $\mathbf{A}$ is governed by the *augmented variational formulation* (cf. [91]):

$$\text{Find } \mathbf{A} \in \mathbf{X}^\alpha_N(\Omega) \text{ such that for all } \mathbf{v} \in \mathbf{X}^\alpha_N(\Omega), \quad a_\alpha(\mathbf{A},\mathbf{v}) = (\mathbf{J},\mathbf{v})_{\mathbf{L}^2(\Omega)}, \tag{5.4.1}$$

where the sesquilinear form

$$a_\alpha(\mathbf{u},\mathbf{v}) := (\mu^{-1}\,\mathbf{curl}\,\mathbf{A}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} + (\mathrm{div}\,\mathbf{A}, \mathrm{div}\,\mathbf{v})_{L^2_\alpha(\Omega)} + \sum_{m=1}^K \langle \mathbf{A}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)} \overline{\langle \mathbf{v}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)}}$$

is continuous and coercive on $\mathbf{X}^\alpha_N(\Omega) \times \mathbf{X}^\alpha_N(\Omega)$. We fall within the framework of the Lax-Milgram theorem. On the other hand, starting from the augmented variational formulation (5.4.1), one can prove that its solution $\mathbf{A}$ is governed by (5.1.15).

Theorem 5.4.1. *Let the data* $\mathbf{J}$ *fulfill assumption (5.1.13). Then, the vector potential* $\mathbf{A}$ *solves (5.1.15) if, and only if,* $\mathbf{A}$ *solves (5.4.1).*

Proof. Let $\mathbf{A}$ be the solution of (5.4.1). We know that $\mathbf{A} \in \mathbf{H}_0(\mathbf{curl},\Omega)$. Next, we take appropriate test functions in (5.4.1). Below, we use a triple orthogonality property, which allows to deal with the three terms defining the form a_α separately.
Note that $\mathbf{v}_\beta = \sum_{k=1}^K \beta_k \nabla q_k$, where $(\beta_k)_{k=1}^K \in \mathbb{C}^K$ and $(q_k)_{k=1}^K$ is given by (5.3.1), belongs to $\mathbf{X}^\alpha_N(\Omega)$. Using it as a test function in (5.4.1), we find that

$$\sum_{m=1}^K \langle \mathbf{A}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)} \overline{\langle \mathbf{v}_\beta\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)}} = 0$$

after integrating by parts the right-hand side, thanks to assumption (5.1.13).
On the other hand, the mapping $(\beta_k)_{k=1}^K \mapsto (\langle \mathbf{v}_\beta\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)})_{m=1}^K$ is onto. This classical result stems from the fact that the capacitance matrix, with entries $((\nabla q_m, \nabla q_k)_{\mathbf{L}^2(\Omega)})_{1\le k,m\le K}$, is invertible (see e.g. [16, cor. 3.3.8]). As a consequence, the fluxes $(\langle \mathbf{A}\cdot\mathbf{n}, 1\rangle_{H^{1/2}(\Gamma_m)})_{m=1}^K$ all vanish.
Going back to (5.4.1), we find that $\mathbf{A}$ is such that, for all $\mathbf{v} \in \mathbf{X}^\alpha_N(\Omega)$, $(\mu^{-1}\,\mathbf{curl}\,\mathbf{A}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} + (\mathrm{div}\,\mathbf{A}, \mathrm{div}\,\mathbf{v})_{L^2_\alpha(\Omega)} = (\mathbf{J},\mathbf{v})_{\mathbf{L}^2(\Omega)}$.

By definition, $\operatorname{div}\mathbf{A} \in L^2_\alpha(\Omega) \subset H^{-1}(\Omega)$, so there exists a unique $q_A \in H^1_0(\Omega)$ such that $-\Delta q_A = \operatorname{div}\mathbf{A}$. The test function $\mathbf{v}_A = \nabla q_A$ can be used above, which now yields $\|\operatorname{div}\mathbf{A}\|^2_{L^2_\alpha(\Omega)} = 0$, after integrating by parts the right-hand side, thanks again to assumption (5.1.13). Therefore, $\mathbf{A}$ is such that, for all $\mathbf{v} \in \mathbf{X}^\alpha_N(\Omega)$, $(\mu^{-1}\,\mathbf{curl}\,\mathbf{A}, \mathbf{curl}\,\mathbf{v})_{\mathbf{L}^2(\Omega)} = (\mathbf{J}, \mathbf{v})_{\mathbf{L}^2(\Omega)}$. Finally, taking $\mathbf{v} \in \mathcal{D}(\Omega)$, we conclude that $\mathbf{curl}(\mu^{-1}\,\mathbf{curl}\,\mathbf{A}) = \mathbf{J}$, and $\mathbf{A}$ solves the magnetic system (5.1.15) as claimed. ⋄

Note that the exact solution $\mathbf{A}$ is recovered exactly for all values of $\alpha \in]\alpha_{\min}, 1[$.

5.4.2 Continuous vector finite elements

We discretize variational formulation (5.4.1) with continuous vector finite elements. We report below the main steps of the analysis, originally carried out in [116]. It starts with the study of singularities, cf. Sections 6 and 7 in *loc. cit.*.

Remark 5.4.1. *Within the taxonomy of [114, §6], one can prove that* $\mathbf{A}$ *contains only singularities of type 1 and type 2, and no singularity of type 3, because* $\mathbf{J} \in \mathbf{L}^2(\Omega)$ *and* $\operatorname{div}\mathbf{A} = 0$ *(see* loc. cit.*, p. 259).*

A solution $\mathbf{A}$ to (5.1.15) can be split into regular and gradient parts as follows:

$$\mathbf{A} = \mathbf{A}_{reg} + \nabla\phi_A, \text{ with} \begin{cases} \mathbf{A}_{reg} \in \mathcal{P}\dot{\mathbf{H}}^{1+\sigma_{Neu}(\mu)}(\Omega) \cap \mathbf{H}_0(\mathbf{curl}, \Omega) \\ \phi_A \in \dot{H}^{1+\sigma_{Dir}}(\Omega) \cap H^1_{\partial\Omega}(\Omega) \end{cases} . \tag{5.4.2}$$

In addition, there exists $C_\alpha > 0$ (independent of $\mathbf{A}$) such that $\|\mathbf{A}\|_{\mathbf{X}^\alpha_N(\Omega)} \le C_\alpha\,(\|\mathbf{A}_{reg}\|_{\mathbf{H}^1(\Omega)} + \|\phi_A\|_{V^2_\alpha(\Omega)})$, where $V^2_\alpha(\Omega)$ is a weighted Sobolev space *à la Nazarov-Plamenevski* (cf. [116, eq. (2.1)]). Next, one uses this regular-gradient splitting (5.4.2) to obtain error estimates; precisely, the error on $\mathbf{A}_{reg}$ in $\mathbf{H}^1(\Omega)$, and the error on ϕ_A in $V^2_\alpha(\Omega)$.

Given $k \ge 1$, we choose the continuous Lagrange elements of order k to define finite dimensional subspaces $(\mathbf{X}_{h,k})_h$ of $\mathbf{X}^\alpha_N(\Omega)$. Given h, one introduces

$$\mathbf{X}_{h,k} = \{\mathbf{v}_h \in \mathbf{H}^1(\Omega) \cap \mathbf{H}_0(\mathbf{curl}, \Omega) \mid \forall \ell \in \mathcal{I}_K, \quad \mathbf{v}_{h|K_\ell} \in \mathbf{P}^k(K_\ell)\}. \tag{5.4.3}$$

The discrete variational formulation of the augmented variational formulation (5.4.1) is then

$$\text{Find } \mathbf{A}^h_\alpha \in \mathbf{X}_{h,k} \text{ such that for all } \mathbf{v}_h \in \mathbf{X}_{h,k}, \quad a_\alpha(\mathbf{A}^h_\alpha, \mathbf{v}_h) = (\mathbf{J}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}. \tag{5.4.4}$$

Observe that the discrete solution $\mathbf{A}^h_\alpha$ depends on both h and α.

Classically, $(\mathbf{X}_{h,k})_h$ exhibits approximability properties for the regular part:

$$\begin{cases} \text{For } \mathtt{s}' = 1 \text{ if } \sigma_{Neu}(\mu) = 1, \text{ and all } \mathtt{s}' \in]0, \sigma_{Neu}(\mu)[\text{ else} \\ \forall \mathbf{v} \in \mathcal{P}\dot{\mathbf{H}}^{1+\sigma_{Neu}(\mu)}(\Omega) \cap \mathbf{H}_0(\mathbf{curl}, \Omega),\ \exists c_{v,\mathtt{s}'},\ \forall h,\ \exists \mathbf{v}_h \in \mathbf{X}_{h,k}, \\ \|\mathbf{v} - \mathbf{v}_h\|_{\mathbf{H}^1(\Omega)} \le c_{v,\mathtt{s}'} h^{\mathtt{s}'}. \end{cases} \tag{5.4.5}$$

In addition, one must have good approximability properties for the gradient part: namely, one needs that there exists a family of spaces $(\Phi_h)_h$ of scalar fields such that

$$\forall h,\ \nabla[\Phi_h] \subset \mathbf{X}_{h,k}, \tag{5.4.6}$$

with[4]

$$\begin{cases} \forall \varepsilon > 0,\ \forall \phi \in \dot{H}^{1+\sigma_{Dir}}(\Omega) \cap H^1_{\partial\Omega}(\Omega),\ \exists c_{\phi,\varepsilon},\ \forall h,\ \exists \phi_h \in \Phi_h, \\ \|\phi - \phi_h\|_{V^2_\alpha(\Omega)} \le c_{\phi,\varepsilon}\, h^{\alpha - \alpha_{\min} - \varepsilon}\ . \end{cases} \tag{5.4.7}$$

A natural question is to check in which case the assumption (5.4.6)-(5.4.7) holds. Below, we report some results in $2D$ and $3D$ domains. The issue was first discussed by Costabel and Dauge in [116]. They observed that, in a $2D$ polygonal domain, one can choose Hsieh-Clough-Tocher finite elements [109] to

[4] Recall that $\alpha_{\min} = 1 - \sigma_{Dir} \in]0, 1/2[$, so the right-hand side of (5.4.7) is equal to $c_{\phi,\varepsilon}\, h^{\sigma_{Dir} + (\alpha-1) - \varepsilon}$.

define the discrete spaces $(\Phi_h)_h$. Indeed, the discrete fields belong to $\mathcal{C}^1(\overline{\Omega})$ and are piecewise P^3 on a subdivision of triangles, so their gradients are naturally in $\mathcal{C}^0(\overline{\Omega})$ and are piecewise $\mathbf{P}^2$. Moreover, one can choose those fields to vanish on the boundary, to enforce the condition on the tangential trace. This ensures that the assumption (5.4.6)-(5.4.7) holds in a $2D$ domain at least for $k = 2$. According to [90, §46], there exist similar finite elements in $2D$ for $k \geq 3$, and also in $3D$: the former can be found in [211], and the latter in [4]. More recent results can be found in [233, 19, 144] and Refs. therein, including the case $k = 1$, and are proved by considering again appropriate subdivisions of tetrahedra and/or triangles.

The numerical analysis is straightforward, because one can apply Céa's lemma. Using the approximability properties above, we conclude that the error estimates below hold.

Theorem 5.4.2 (`[A4, Theorem 14]`). *For $\alpha \in]\alpha_{\min}, 1[$, let $\tau_\alpha = \min(\sigma_{Neu}(\mu), \sigma_{Dir} + (\alpha - 1))$. For $h > 0$, let $\mathbf{A}_\alpha^h$ be the solution to (5.4.4). Then, for all $\varepsilon > 0$, there exists $c_{A,\varepsilon}$ such that for all h:*

$$\|\mathbf{A} - \mathbf{A}_\alpha^h\|_{\mathbf{X}_N^\alpha(\Omega)} \leq c_{A,\varepsilon} h^{\tau_\alpha - \varepsilon}, \tag{5.4.8}$$

Going back to the magnetic field $\mathbf{B}_J$, that it is now guaranteed that

$$\|\mathbf{B}_J - \mathbf{curl}\, \mathbf{A}_\alpha^h\|_{\mathbf{H}(\mathrm{div},\Omega)} \leq c_{J,\varepsilon} h^{\tau_\alpha - \varepsilon}.$$

Within the framework of the WRM, we can take any $\alpha \in]\alpha_{\min}, 1[$ so, if one picks α "close to 1", the a priori error is almost like $O(h^{\tau_{\max}})$, where $\tau_{\max} = \min(\sigma_{Neu}(\mu), \sigma_{Dir})$.

We note that the linear system equivalent to (5.4.4) is expected to be easy to solve because the form a_α is coercive. In $2D$ and $3D$ domains, one can find numerical experiments [A8, A1, 208, 209] to support the claims, and to show the robustness of the approach. In particular, in [208, 209], a very coarse approximation of the weight w_α is chosen, namely a discrete weight which is equal to 0 in simplices located near the reentrant edges, and equal to 1 in all other simplices.

5.5 Eigenvalue problem, continuous vector finite elements [A1]

Let $\lambda = \omega^2/c^2$. The eigenpair $(\mathbf{E}, \lambda)$ solving the electric eigenproblem (5.1.18) is a solution of:

$$\text{Find } (\mathbf{E}, \lambda) \in \mathbf{X}_N^\alpha(\Omega) \times \mathbb{R}^+ \text{ such that } \mathbf{curl}\, \mathbf{curl}\, \mathbf{E} = \lambda \mathbf{E}. \tag{5.5.1}$$

Consider first the following augmented variational formulation:

$$\text{Find } (\mathbf{E}, \lambda) \in \mathbf{X}_N^\alpha(\Omega) \times \mathbb{R}^+ \text{ such that for all } \mathbf{v} \in \mathbf{X}_N^\alpha(\Omega), \quad (\mathbf{E}, \mathbf{v})_{\mathbf{X}_N^\alpha(\Omega)} = \lambda (\mathbf{E}, \mathbf{v})_{\mathbf{L}^2(\Omega)}.$$

Unfortunately the fact that $(\mathbf{E}, \lambda)$ is a solution to the above problem does not guarantee that it is an eigenpair of the original problem. Indeed, there exists solutions, called *spurious eigenpairs* which are not divergence-free. To address this problem, Costabel and Dauge [116, 117] proposed to consider the *parametrized eigenproblem*:

$$\begin{cases} \text{Find } (\mathbf{E}_s, \lambda_s) \in \mathbf{X}_N^\alpha(\Omega) \times \mathbb{R}^+ \text{ such that for all } \mathbf{v} \in \mathbf{X}_N^\alpha(\Omega) \\ \quad (\mathbf{curl}\, \mathbf{E}, \mathbf{curl}\, \mathbf{v})_{\mathbf{L}^2(\Omega)} + s\, (\mathrm{div}\, \mathbf{E}, \mathrm{div}\, \mathbf{v})_{\mathbf{L}^2(\Omega)} = \lambda_s (\mathbf{E}, \mathbf{v})_{\mathbf{L}^2(\Omega)} \end{cases},$$

where $s > 0$ is a parameter. The equivalence with the original problem can be restored. If one let s vary, the true eigenpairs, which are divergence free, will be independent of s, whereas the spurious solutions still vary with s (their divergence doesn't vanish).

A second approach, following [91], consists in taking into account the constraint of the divergence of the field with a Lagrange multiplier. We recover thus the equivalence with the original problem. The

mixed eigenproblem to be solved reads:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{E},p,\lambda)\in \mathbf{X}_N^\alpha(\Omega)\times L^2_\alpha(\Omega)\times\mathbb{R}^+ \text{ such that for all } (\mathbf{v},q)\in \mathbf{X}_N^\alpha(\Omega)\times L^2_\alpha(\Omega)\\ \begin{array}{rcll} (\mathbf{E},\mathbf{v})_{\mathbf{X}_N^\alpha(\Omega)}+(p,\operatorname{div}\mathbf{v})_{L^2_\alpha(\Omega)} &=& \lambda(\mathbf{E},\mathbf{v})_{\mathbf{L}^2(\Omega)} & \text{(i)}\\ (q,\operatorname{div}\mathbf{E})_{L^2_\alpha(\Omega)} &=& 0 & \text{(ii)} \end{array} \end{array}\right. . \tag{5.5.2}$$

Theorem 5.5.1. *Problems* (5.5.1) *and* (5.5.2) *are equivalent.*

Proof. Let $(\mathbf{E},p,\lambda)$ be a solution to (5.5.2). According to (5.5.2)-(ii), $\operatorname{div}\mathbf{E}=0$. Let $\phi_p\in H^1_0(\Omega)$ be the unique solution to $\Delta\phi_p=p$ (it is a well-posed problem since $L^2_\alpha(\Omega)\subset H^{-1}(\Omega)$). Let $\mathbf{v}_p=\nabla\phi_p$. By construction, $\mathbf{curl}\,\mathbf{v}_p=0$, $\mathbf{v}\times\mathbf{n}_{|\partial\Omega}=0$ and $\operatorname{div}\mathbf{v}_p=p$, hence $\mathbf{v}_p\in\mathbf{X}_N^\alpha(\Omega)$. If we choose $\mathbf{v}=\mathbf{v}_p$ in (5.5.2)-(i), we get: $\|p\|^2_{L^2_\alpha(\Omega)}=\lambda(\mathbf{E},\nabla\phi_p)_{\mathbf{L}^2(\Omega)}=0$ by integration by parts. Therefore, any solution of (5.5.2) is such that $\mathbf{E}\in\mathbf{X}_N^\alpha(\Omega)$ and $p=0$. ⋄

5.5.1 Numerical analysis

To carry out the numerical analysis of problem 5.5.2 discretized with $\mathbf{P}^{k+1}\times P^k$ finite elements (cf. problem (5.5.11) below), we study the resolution of the static problem posed in a mixed augmented variational formulation. Let us recall the static problem:

$$\text{Find } \tilde{\mathbf{E}}\in\mathbf{H}_0(\mathbf{curl},\Omega)\cap\mathbf{H}(\operatorname{div}0,\Omega) \text{ such that } \mathbf{curl}\,\mathbf{curl}\,\tilde{\mathbf{E}}=\tilde{\mathbf{J}}, \text{ with } \tilde{\mathbf{J}}\in\mathbf{H}(\operatorname{div}0,\Omega). \tag{5.5.3}$$

Recall first that we can discretize problem (5.5.3) with the WRM with $\mathbf{P}^{k+1}$ finite elements. The discrete problem reads:

$$\text{Find } \tilde{\mathbf{E}}_h\in\mathbf{X}_{h,k+1}\,|\,\forall\mathbf{v}_h\in\mathbf{X}_{h,k+1},\quad (\tilde{\mathbf{E}}_h,\mathbf{v}_h)_{\mathbf{X}_N^\alpha(\Omega)}=(\tilde{\mathbf{J}},\mathbf{v}_h)_{\mathbf{L}^2(\Omega)}$$

According to theorem 5.4.2, the following estimate holds:

$$\|\tilde{\mathbf{E}}-\tilde{\mathbf{E}}_h\|_{\mathbf{X}_N^\alpha(\Omega)}\le C_\mu\,h^\mu\,\|\tilde{\mathbf{J}}\|_{\mathbf{L}^2(\Omega)},\quad \forall\mu\in]0,\tau_\alpha[, \tag{5.5.4}$$

where $C_\mu>0$ does not depend on the mesh size.
The mixed augmented variational formulation related to (5.5.3) reads:

$$\left\{\begin{array}{l} \text{Find } (\tilde{\mathbf{E}},\tilde{p})\in \mathbf{X}_N^\alpha(\Omega)\times L^2_\alpha(\Omega) \text{ such that for all } (\mathbf{v},q)\in \mathbf{X}_N^\alpha(\Omega)\times L^2_\alpha(\Omega)\\ \begin{array}{rcll} (\tilde{\mathbf{E}},\mathbf{v})_{\mathbf{X}_N^\alpha(\Omega)}+(\tilde{p},\operatorname{div}\mathbf{v})_{L^2_\alpha(\Omega)} &=& (\tilde{\mathbf{J}},\mathbf{v})_{\mathbf{L}^2(\Omega)} & \text{(i)}\\ (q,\operatorname{div}\tilde{\mathbf{E}})_{L^2_\alpha(\Omega)} &=& 0 & \text{(ii)} \end{array} \end{array}\right. . \tag{5.5.5}$$

We set: $\mathbf{K}_N^\alpha(\Omega):=\mathbf{X}_N^\alpha(\Omega)\cap\mathbf{H}(\operatorname{div}0,\Omega)$. We want to apply abstract theory recalled in §4.5 with $V=\mathbf{X}_N^\alpha(\Omega)$, $Q=L^2_\alpha(\Omega)$ $a(u,v)=(u,v)_{\mathbf{X}_N^\alpha(\Omega)}$, $b(v,q)=(\operatorname{div}v,q)_{L^2_\alpha(\Omega)}$, $\mathbb{K}=\mathbf{K}_N^\alpha(\Omega)$. Applying theorem 4.5.1, we have the

Theorem 5.5.2. *Problem* (5.5.5) *is well-posed.*

Proof. The bilinear form $(\mathbf{w},\mathbf{v})_{\mathbf{X}_N^\alpha(\Omega)}$ is obviously coercive on $\mathbf{K}_N^\alpha(\Omega)$. Let $q\in L^2_\alpha(\Omega)$ and $\phi_q\in H^1_0(\Omega)$ such that $\Delta\phi_q=q$. Then $\mathbf{v}_q=\nabla\,\phi_q\in\mathbf{X}_N^\alpha(\Omega)$ (cf. the proof of theorem 5.5.1), and $(q,\operatorname{div}\mathbf{v}_q)_{L^2_\alpha(\Omega)}=\|q\|_{L^2_\alpha(\Omega)}\|\mathbf{v}_q\|_{\mathbf{X}_N^\alpha(\Omega)}$: the inf-sup condition (4.5.2) holds with $\beta=1$. ⋄

We introduce two functional spaces: $V^1_\alpha(\Omega):=\{v\in L^1_{loc}(\Omega)\,|\,w_{\alpha-1}v\in L^2(\Omega),\,w_\alpha\nabla v\in\mathbf{L}^2(\Omega)\}$, $\overset{\circ}{V}{}^1_\alpha(\Omega):=$ closure of $\mathcal{D}(\Omega)$ in $V^1_\alpha(\Omega)$. We know from [116] that $\overset{\circ}{V}{}^1_\alpha(\Omega)=\{v\in V^1_\alpha(\Omega)\,|\,v_{|\partial\Omega}=0\}$ and that on $\overset{\circ}{V}{}^1_\alpha(\Omega)$, the inner norm and the semi-norm are equivalent norms: $\|v\|_{\overset{\circ}{V}{}^1_\alpha(\Omega)}=\|\nabla(w_\alpha v)\|_{\mathbf{L}^2(\Omega)}$. This implies the orthogonal decomposition

$$\mathbf{L}^2(\Omega)=\mathbf{H}(\operatorname{div}0,\Omega)\overset{\perp}{\oplus}\nabla(w_\alpha\overset{\circ}{V}{}^1_\alpha(\Omega)) \tag{5.5.6}$$

In the absence of weight, $\overset{\circ}{V}{}^1_\alpha(\Omega)=H^1_0(\Omega)$. Let $Q_\alpha(\Omega)=\{q\in L^2_\alpha(\Omega)\,|\,w_\alpha q\in\overset{\circ}{V}{}^1_\alpha(\Omega)\}$, endowed with the

norm: $\|q\|_{Q_\alpha(\Omega)} = \|\nabla(w_{2\gamma} q)\|_{\mathbf{L}^2(\Omega)}$. We have the

Theorem 5.5.3 (`[A1, Theorem 4.5]`). *Let $(\tilde{\mathbf{E}}, \tilde{p})$ be the solution to* (5.5.5). *On the one hand,* $\operatorname{div} \tilde{\mathbf{E}} = 0$, *and* $\mathbf{curl}\,\mathbf{curl}\,\tilde{\mathbf{E}} \in \mathbf{L}^2(\Omega)$. *On the other hand,* $\tilde{p} \in Q_\alpha(\Omega)$.

Proof. According to (5.5.5)-(ii), $\operatorname{div} \tilde{\mathbf{E}} = 0$. According to (5.5.6), $\exists!(\mathbf{w}, q) \in \mathbf{H}(\operatorname{div} 0, \Omega) \times \overset{\circ}{V}{}^1_\alpha(\Omega)$ s. t. $\tilde{\mathbf{J}} = \mathbf{w} + \nabla(w_\alpha q)$. Then $w_\alpha p + q \in L^2(\Omega)$ and we can solve the auxiliary problem:

$$\text{Find } z \in \overset{\circ}{V}{}^1_\alpha(\Omega) \,|\, \forall z' \in \overset{\circ}{V}{}^1_\alpha(\Omega), \quad (\nabla(w_\alpha z), \nabla(w_\alpha z'))_{\mathbf{L}^2(\Omega)} = -(w_\alpha p + q, z')_{L^2\Omega)}.$$

By construction, $w_\alpha \Delta(w_\alpha z) = w_\alpha p + q$. Let $\mathbf{v} = \nabla(w_\alpha z) \in \mathbf{X}^\alpha_N(\Omega)$. It holds: $\mathbf{curl}\,\mathbf{v} = 0$ and $w_\alpha \operatorname{div} \mathbf{v} = w_\alpha p + q$. Using this test-field in (5.5.5)-(i) yields: $(w_\alpha p, w_\alpha p + q)_{L^2(\Omega)} = (\tilde{\mathbf{J}}, \nabla(w_\alpha z))_{\mathbf{L}^2(\Omega)} \overset{\perp}{=} (\nabla(w_\alpha q), \nabla(w_\alpha z))_{\mathbf{L}^2(\Omega)} \overset{z' = q}{=} -(w_\alpha p + q, q)_{L^2(\Omega)}$. In other words, $\|w_\alpha p + q\|^2_{L^2(\Omega)} = 0$, i.e. $w_\alpha p = -q$ in $L^2(\Omega)$: this proves the improved regularity of p.
Now, let us take a test-field $\mathbf{v} \in \mathcal{D}(\Omega)$ in (5.5.5)-(i), with $\tilde{\mathbf{J}}$ replaced by $\mathbf{w} + \nabla(w_\alpha q)$:

$$(\mathbf{curl}\,\mathbf{curl}\,\tilde{\mathbf{E}} - \mathbf{w}, \mathbf{v})_{\mathbf{L}^2(\Omega)} = (\nabla(w_\alpha q), v)_{\mathbf{L}^2(\Omega)} - (w_\alpha p, w_\alpha \operatorname{div} \mathbf{v})_{L^2(\Omega)} = -(q + w_\alpha p, w_\alpha \operatorname{div} \mathbf{v})_{L^2(\Omega)} = 0.$$

Hence, $\mathbf{curl}\,\mathbf{curl}\,\tilde{\mathbf{E}} = \mathbf{w}$, so the conclusion follows. ◇

Corollary 5.5.1 (`[A1, cor. 4.1]`). *One can characterize the spaces V_0 and Q_0 defined in* (4.5.13) *completely. After some elementary computations, one reaches $V_0 = \mathbf{V}_{\mathbf{E}}(\Omega) := \{\mathbf{v} \in \mathbf{K}^\alpha_N \,|\, \mathbf{curl}\,\mathbf{curl}\,\mathbf{v} \in \mathbf{L}^2(\Omega)\}$ and $Q_0 = Q_\alpha(\Omega)$. The space $\mathbf{V}_{\mathbf{E}}(\Omega)$ is endowed with the norm:*

$$\|\mathbf{v}\|_{\mathbf{V}_{\mathbf{E}}(\Omega)} = \left(\|\mathbf{curl}\,\mathbf{v}\|^2_{\mathbf{L}^2(\Omega)} + \|\mathbf{curl}\,\mathbf{curl}\,\mathbf{v}\|^2_{\mathbf{L}^2(\Omega)} \right)^{\frac{1}{2}}. \tag{5.5.7}$$

Consider the (self-adjoint) operator $\mathsf{T} : (\mathbf{X}^\alpha_N(\Omega))' \to \mathbf{X}^\alpha_N(\Omega)$ such that $\mathsf{T}(\tilde{\mathbf{J}}) = \tilde{\mathbf{E}}$, where $(\tilde{\mathbf{E}}, \tilde{p})$ is the solution to problem (5.5.5). The space $V = \mathbf{X}^\alpha_N(\Omega)$ is dense in $L = \mathbf{L}^2(\Omega)$. The last corollary allows us to prove the compactness result (E):

Corollary 5.5.2 (`[A1, cor. 4.2]`). *The restriction of operator T from $\mathbf{L}^2(\Omega)$ to $\mathbf{V}_{\mathbf{E}}(\Omega)$ is compact.*

Proof. Consider a bounded sequence $(\tilde{\mathbf{J}}_n)_n \in \mathbf{L}^2(\Omega)$. The corresponding solutions $(\tilde{\mathbf{E}}_n, \tilde{p}_n)$ are bounded in $\mathbf{V}_{\mathbf{E}}(\Omega) \times Q_\alpha(\Omega)$. In particular, thanks to the identification (5.5.7), the sequence $(\mathbf{F}_n)_n$, with $\mathbf{F}_n = \mathbf{curl}\,\tilde{\mathbf{E}}_n$, is bounded in the space $\mathbf{X}_T(\Omega) := \mathbf{H}_0(\operatorname{div}, \Omega) \cap \mathbf{H}(\mathbf{curl}, \Omega)$. According to the fundamental result of Weber [244], $\mathbf{X}_T(\Omega)$ is compactly embedded in $\mathbf{L}^2(\Omega)$: there exists $\mathbf{F} \in \mathbf{L}^2(\Omega)$ s. t. $\lim_{n\to\infty} \mathbf{F}_n = \mathbf{F} \in \mathbf{L}^2(\Omega)$ (we identify the subsequence of $(\mathbf{F}_n)_n$ with $(\mathbf{F}_n)_n$ itself). Obviously, the sequence $(\tilde{\mathbf{E}}_n)_n$ is bounded in $\mathbf{X}_N(\Omega)$: according to the second fundamental result of Weber [244] on vector fields with vanishing tangential trace, $\mathbf{X}_N(\Omega)$ is also compactly embedded in $\mathbf{L}^2(\Omega)$. Therefore, there exists $\tilde{\mathbf{E}} \in \mathbf{L}^2(\Omega)$ such that $\lim_{n\to\infty} \tilde{\mathbf{E}}_n = \tilde{\mathbf{E}} \in \mathbf{L}^2(\Omega)$. Since $\operatorname{div} \tilde{\mathbf{E}}_n = 0$, for all n, one has $\operatorname{div} \tilde{\mathbf{E}} = 0$ thanks to the uniqueness of the limit. Moreover, $\mathbf{F}_n = \mathbf{curl}\,\tilde{\mathbf{E}}_n$ converges to $\mathbf{F}$ in $\mathbf{L}^2(\Omega)$, and by identification $\mathbf{curl}\,\tilde{\mathbf{E}} = \mathbf{F}$. Then, the subsequence $(\tilde{\mathbf{E}}_n)_n$ converges to $\tilde{\mathbf{E}}$ in $\mathbf{H}(\mathbf{curl}, \Omega)$, so that $\mathbf{E} \in \mathbf{H}_0(\mathbf{curl}, \Omega)$. Finally one concludes first that $\tilde{\mathbf{E}}$ belongs to $\mathbf{X}^\alpha_N(\Omega)$, and second that the subsequence $(\tilde{\mathbf{E}}_n)_n$ converges to $\tilde{\mathbf{E}}$ in $\mathbf{X}^\alpha_N(\Omega)$, as expected. ◇

Compare to §5.4, we want here to approximate the fields by an element of the discrete kernel. The obvious choice of the Taylor-Hood Finite Element [239] is ill-advised in our case. A condition like the discrete Lagrange multiplier vanishes on tetrahedra that neighbour the reentrant edges and corners must be imposed. We can discretize (5.5.5) and (5.5.2) with the Zero Near Singularity $\mathbf{P}^{k+1} \times P^k$ Finite Element of Ciarlet Jr. and Hechme [100]: to discretize the electric field, we consider $\mathbf{X}_{h,k+1}$ defined in (5.4.3). Let $E_h = \bigcup_{T \in \mathcal{T}_h, \partial T \cap \mathcal{E} \neq 0} T$. To discretize the Lagrange multiplier, we consider

$$M_h := \{q_h \in \mathcal{C}^0(\overline{\Omega}) \,|\, \forall \ell \in \mathcal{I}_K,\, w_{2\alpha} q_{h|K_\ell} \in P^k(K_\ell) \text{ and } q_{h|E_h} = 0\},$$

Moreover, given $q_h \in M_h$, letting $\overline{q}_h = w_{2\alpha} q_h$, we have: $(q_h, \operatorname{div} \mathbf{v}_h)_{L^2_\alpha(\Omega)} = (\overline{q}_h, \operatorname{div} \mathbf{v}_h)_{L^2\Omega)} = 0$ for all $\mathbf{v}_h \in \mathbf{X}^\alpha_N(\Omega)$.

The discrete approximation of (5.5.5) reads:
Find $(\tilde{\mathbf{E}}_h, \tilde{p}_h) \in \mathbf{X}_{h,k+1} \times M_h$ such that for all $(\mathbf{v}_h, q_h) \in \mathbf{X}_{h,k+1} \times M_h$,

$$\left\{ \begin{array}{rcll} (\tilde{\mathbf{E}}_h, \mathbf{v}_h)_{\mathbf{X}^\alpha_N(\Omega)} + (\tilde{p}_h, \operatorname{div} \mathbf{v}_h)_{L^2_\alpha(\Omega)} & = & (\tilde{\mathbf{J}}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} & \text{(i)} \\ (q_h, \operatorname{div} \tilde{\mathbf{E}}_h)_{L^2_\alpha(\Omega)} & = & 0 & \text{(ii)} \end{array} \right. . \tag{5.5.8}$$

Applying theorem 4.5.1 with estimate (5.5.4), we obtain the error estimate for $\tilde{\mathbf{E}}_h$ approximated with (5.5.8):

$$\|\tilde{\mathbf{E}} - \tilde{\mathbf{E}}_h\|_{\mathbf{X}_N^\alpha(\Omega)} \leq C_\mu \left(1 + \frac{1}{\beta(h)}\right) h^\mu \|\tilde{\mathbf{J}}\|_{\mathbf{L}^2(\Omega)}, \quad \forall \mu \in]0, \tau_\alpha[, \tag{5.5.9}$$

where $\beta(h)$ is the constant of the discrete inf-sup condition (4.5.5) on $\mathbf{X}_{h,k+1} \times M_h$, with $b(\mathbf{v}_h, q_h) = (\operatorname{div} \mathbf{v}_h, q_h)_{L^2_\alpha(\Omega)}$. From now on, we suppose that $\beta(h)$ is bounded from below. Noticing that $\|\tilde{\mathbf{J}}\|_{\mathbf{L}^2(\Omega)} \leq \|\tilde{\mathbf{E}}\|_{\mathbf{V}_\mathbf{E}(\Omega)}$, this results allows to obtain the bound $r_2(h) \leq C_\mu(1 + \beta(h)^{-1})h^\mu$, for $\mu \in]0, \tau[$ giving the *strong approximation property* (def. 4.5.2).

Next, we establish an approximability property for the Lagrange multiplier.

Proposition 5.5.1 (`[A1, Proposition 4.1]`). *There exists $C_{app} > 0$ such that for all $q \in Q_\alpha(\Omega)$,*

$$\inf_{q_h \in M_h} \|q - q_h\|_{L^2_\alpha(\Omega)} \leq C_{app} h^{1-\alpha} \|q\|_{Q_\alpha(\Omega)}. \tag{5.5.10}$$

Proof. We follow the proof of [116, Proposition 8.3], with Clément type interpolation operators since functions of $Q_\alpha(\Omega)$ are not, in general, continuous. Note that $q \in Q_\alpha(\Omega)$ mean that $w_{2\alpha-1}q \in L^2(\Omega)$ and $w_{2\alpha}\nabla q \in \mathbf{L}^2(\Omega)$. For all $\ell \in \mathcal{I}_K$, we call $\tilde{K}_\ell = \{K_{\ell'},\, \ell' \in \mathcal{I}_K,\, | K_\ell \cap K_{\ell'} \neq \emptyset\}$. Let $V_h = \{q_j \in \mathcal{C}^0(\overline{\Omega}) \,|\, \forall \ell \in \mathcal{I}_K,\, q_{h|K_\ell} \in P^1(K_\ell)\}$. There exists a projection $\pi_h : L^2_{loc} \to V_h$ such that $\|\tilde{q} - \pi_h(\tilde{q})\|_{L^2(K_\ell)} \lesssim h\, |q|_{H^1(\tilde{K}_\ell)}$ [231]. We construct a cut-off function χ_h such that (recall that d is the distance from the set of reentrant edges):

$$0 \leq \chi_h \leq 1, \quad \chi_h(\mathbf{x}) = 1 \text{ if } d(\mathbf{x}) \leq 4\,h, \quad \chi_h(\mathbf{x}) = 0 \text{ if } d(\mathbf{x}) \leq 8\,h.$$

For any $\ell \in \mathcal{I}_K$, we suppose that if $\chi_h(\mathbf{x}) = 0$, $\forall \mathbf{x} \in K_\ell$, then $d_{\tilde{K}_\ell} \geq h$. Given $q \in Q_\alpha(\Omega)$, we set $q_h = w_{-2\alpha}\pi_h(1 - \chi_h)w_{2\alpha}q$, and we prove (5.5.10) letting $\|q\|^2_{Q_\alpha(\Omega)} \simeq \|q\|^2_{L^2_{2\alpha-1}(\Omega)} + \|\nabla q\|^2_{L^2_{2\alpha}(\Omega)}$. We have that $q_{h|E_h} = 0$, $q_h \in M_h$ and:

$$\|q - q_h\|^2_{L^2_\alpha(\Omega)} \leq 2\,\|\chi_h q\|^2_{L^2_\alpha(\Omega)} + 2\,\|(1 - \chi_h)q - q_h\|^2_{L^2_\alpha(\Omega)} = 2\,(I + II).$$

Let $\tilde{\mathcal{I}}_h = \{\ell \in \mathcal{I}_K \,|\, d_{K_\ell} < 8h\}$. We compute:

$$I = \sum_{\ell \in \tilde{\mathcal{I}}_h} \int_{K_\ell} w_{2\alpha} |\chi_h q|^2 \leq \sum_{\ell \in \tilde{\mathcal{I}}_h} \int_{K_\ell} w_{2\alpha} |q|^2 \lesssim \sum_{\ell \in \tilde{\mathcal{I}}_h} h^{2-2\alpha} \int_{K_\ell} w_{4\alpha-2}\, |q|^2 \lesssim h^{2(1-\alpha)} \|q\|^2_{L^2_{2\alpha-1}(\Omega)}.$$

Letting $\tilde{q} = (1 - \chi_h)w_{2\alpha}q$. We have:

$$II = \|w_{-\alpha}(w_{2\alpha}(1 - \chi_h)q - \pi_h(1 - \chi_h)w_{2\alpha}q)\|^2_{L^2(\Omega)} = \|w_{-\alpha}(\tilde{q} - \pi_h\,\tilde{q})\|^2_{L^2(\Omega)} = \sum_{\ell \in \mathcal{I}_K} \|w_{-\alpha}(\tilde{q} - \pi_h\,\tilde{q})\|^2_{L^2(K_\ell)}.$$

In the above summation, we exclude tetrahedra of E_h. For $K_\ell \in \mathcal{T}_h \backslash E_h$, we have $\|w_{-\alpha}\|_{L^\infty}(K_\ell) \lesssim h^{-\alpha}$. Using the rule $\|fg\|_{L^2} \leq \|f\|_{L^\infty}\|g\|_{L^2}$, it holds:

$$II \lesssim h^{-2\alpha} \sum_{K_\ell \in \mathcal{T}_h \backslash E_h} \|\tilde{q} - \pi_h\,\tilde{q}\|^2_{L^2(K_\ell)} \lesssim h^{2-2\alpha} \sum_{K_\ell \in \mathcal{T}_h \backslash E_h} |\tilde{q}|^2_{H^1(\tilde{K}_\ell)}.$$

Next, we write $\nabla\,\tilde{q} = (1 - \chi_h)\,\nabla(w_{2\alpha}q) + w_{2\alpha}q\nabla(1 - \chi_h)$.
For the first term, one has:

$$\|(1 - \chi_h)\,\nabla(w_{2\alpha}q)\|_{L^2(\tilde{K}_\ell)} \leq \|1 - \chi_h\|_{L^\infty(\tilde{K}_\ell)}\,\|\nabla(w_{2\alpha}q)\|_{L^2(\tilde{K}_\ell)} \leq \|\nabla(w_{2\alpha}q)\|_{L^2(\tilde{K}_\ell)}.$$

For the second term, the support of $\nabla(1 - \chi_h)$ is included in $\{\mathbf{x} \,|\, 4h \leq d(\mathbf{x}) \leq 8h\}$, and

$$\|\nabla(1 - \chi_h)\|_{L^\infty} \lesssim h^{-1}.$$

Also, in this region, one has $\|w_1\|_{L^\infty} \lesssim h$. Hence:

$$\|w_{2\alpha}q\nabla(1 - \chi_h)\|_{L^2(\tilde{K}_\ell)} = \|w_{2\alpha-1}qw_1\nabla(1 - \chi_h)\|_{L^2(\tilde{K}_\ell)} \lesssim \|w_{2\alpha-1}q\|_{L^2(\tilde{K}_\ell)} = \|q\|_{L^2_{2\alpha-1}(\tilde{K}_\ell)}.$$

We obtain then the following estimate for II:

$$II \lesssim h^{2(1-\alpha)} \sum_{\ell \in \mathcal{I}_K} \left(\|\nabla(w_{2\alpha}q)\|^2_{L^2(\tilde{K}_\ell)} + \|q\|^2_{L^2_{2\alpha-1}(\tilde{K}_\ell)}\right) \lesssim h^{2(1-\alpha)} \left(\|\nabla(w_{2\alpha}q)\|^2_{L^2(\Omega)} + \|q\|^2_{L^2_{2\alpha-1}(\Omega)}\right).$$

$\diamond$

As a corollary, we have the *weak approximability property* of $Q_\alpha(\Omega)$ (def. 4.5.1):

Corollary 5.5.3. *Let* $\mathbf{K}_{h,k+1} := \{\mathbf{v}_h \in \mathbf{X}_{h,k+1} \,|\, \forall q_h \in M_h,\ (\operatorname{div} \mathbf{v}_h, q_h)_{L^2_\alpha(\Omega)} = 0\}$. *There exists* $r_1 : \mathbb{R}^+ \to \mathbb{R}^+$ *such that* $\lim_{h\to 0^+} r_1(h) = 0$ *and*

$$\sup_{\mathbf{v}_h \in \mathbf{K}_{h,k+1}} \frac{(\operatorname{div} \mathbf{v}_h, q)}{\|\mathbf{v}_h\|_{\mathbf{X}^\alpha_N(\Omega)}} \leq r_1(h) \|q_h\|_{Q_\alpha}.$$

Proof. Consider $(\mathbf{v}_h, q) \in \mathbf{K}_{h,k+1} \times Q_\alpha(\Omega)$. Noticing that $(\operatorname{div} \mathbf{v}_h, q)_{L^2_\alpha(\Omega)} = (\operatorname{div} \mathbf{v}_h, q - q_h)_{L^2_\alpha(\Omega)}$, $\forall q_h \in M_h$, it holds:

$$|(\operatorname{div} \mathbf{v}_h, q)_{L^2_\alpha(\Omega)}| \leq \|\operatorname{div} \mathbf{v}_h\|_{L^2_\alpha(\Omega)} \|q - q_h\|_{L^2_\alpha(\Omega)} \leq \|\mathbf{v}_h\|_{\mathbf{X}^\alpha_N(\Omega)} \|q - q_h\|_{L^2_\alpha(\Omega)}, \quad \forall q_h \in M_h.$$

We obtain the result with $r_1(h) = C_{app} h^{1-\alpha}$. $\diamond$

The discrete approximation of (5.5.2) reads:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{E}_h, p_h, \lambda_h) \in \mathbf{X}_{h,k+1} \times M_h \text{ such that for all } (\mathbf{v}_h, q_h) \in \mathbf{X}_{h,k+1} \times M_h \\ (\mathbf{E}_h, \mathbf{v}_h)_{\mathbf{X}^\alpha_N(\Omega)} + (p_h, \operatorname{div} \mathbf{v}_h)_{L^2_\alpha(\Omega)} = \lambda_h (\mathbf{E}_h, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} \quad \text{(i)} \\ (q_h, \operatorname{div} \mathbf{E}_h)_{L^2_\alpha(\Omega)} = 0 \quad \text{(ii)} \end{array} \right. . \tag{5.5.11}$$

We conclude with the error estimates. From (5.5.4), we know that the bound in (4.5.18) behaves like $\varepsilon(h) = C_\mu\, h^\mu$, for $\mu \in]0, \tau_\alpha[$ (it can be better-smaller-provided the eigenfield is smooth). Therefore, we have the following error bounds for the discretization of the electric eigenproblem (5.5.1):

- Approximation of the eigenvalue: $|\lambda - \lambda_h| \leq (C_\mu)^2\, h^{2\mu}$,
- Gap: $\hat{\delta}(E_\lambda, E_{\lambda_h}) < C_\mu\, h^\mu$.

5.5.2 Numerical illustration

We consider the benchmark of Dauge [125]. Let Ω be the L-shape domain with straight sides and corners in $(0,0)$, $(1,0)$, $(1,1)$, $(-1,1)$, $(-1,-1)$, $(0,-1)$. It possesses a reentrant corner located at the origin. It is reported that the first eigenvector has the strong unbounded singularity, which means that it does not belong to $\mathbf{H}^1(\Omega)$. According to the theory, the weight w_α can be chosen as $w_\alpha = r^\alpha$, with r the distance to the origin, and $\alpha \in]1/3, 1[$. We compute the solution using a value of the parameter α set to $\alpha = 0.95$, on a series of quasi-uniform meshes. To discretize the problem, we consider the $\mathbf{P}^2 \times P^1$ Zero Near Singularity finite element pair, cf. [100]. The meshes are respectively made of 738, 2 952 and 11 808 triangles, with 410, 1 557 and 6 065 vertices. The corresponding meshsizes are $h = 7.84 \times 10^{-2}$, $h = 3.92 \times 10^{-2}$ and $h = 1.96 \times 10^{-2}$. The computation of the first five eigenvalues $(\lambda_i, h)_{i=1}^5$ is carried out with the help of Matlab. The relative errors $r_{i,h} = |\lambda_{i,h} - \lambda_i|/\lambda_i$ are reported in Table 5.1.

We note that on the coarsest mesh, all eigenvalues are already very well-approximated, with the exception of the first one, for which convergence is in addition slower.

To improve the convergence towards the first eigenvalue, we then make some computations, on a series of suitably graded meshes, cf. Figure 5.1, with the same value of the parameter α, that is $\alpha = 0.95$. See [69] for a rigorous study of the benefit of graded meshes for the computation of Maxwell eigenvalues. It is important to note that the number of triangles and vertices are similar, since the graded meshes are respectively made of 648, 2 664 and 10 728 triangles, with 362, 1 410 and 5 522 vertices. The results on these graded meshes are reported in Table 5.2. Notice that the relative erros $r_{3,h}$ and $r_{5,h}$ are better on the first graded mesh. We need further investigation. This may be explained by the fact that the graded meshes are not well suited to the corresponding solutions. To complete this series of experiments, let us emphasize the crucial importance of the weight w_α . Without the weight, it is expected that the

eigenvalue(s) corresponding to singular eigenvector(s) are not captured numerically (in our case, the first one).

h	$r_{1,h}$	$r_{2,h}$	$r_{3,h}$	$r_{4,h}$	$r_{5,h}$
Uniform 1	1.3×10^{-2}	3.3×10^{-4}	9.4×10^{-5}	1.1×10^{-4}	9.9×10^{-3}
Uniform 2	8.0×10^{-3}	6.2×10^{-5}	2.3×10^{-5}	2.5×10^{-5}	1.3×10^{-5}
Uniform 3	4.4×10^{-3}	1.2×10^{-5}	5.5×10^{-6}	6.2×10^{-6}	5.3×10^{-6}

Table 5.1: Relative errors on quasi-uniform meshes with $\alpha = 0.95$.

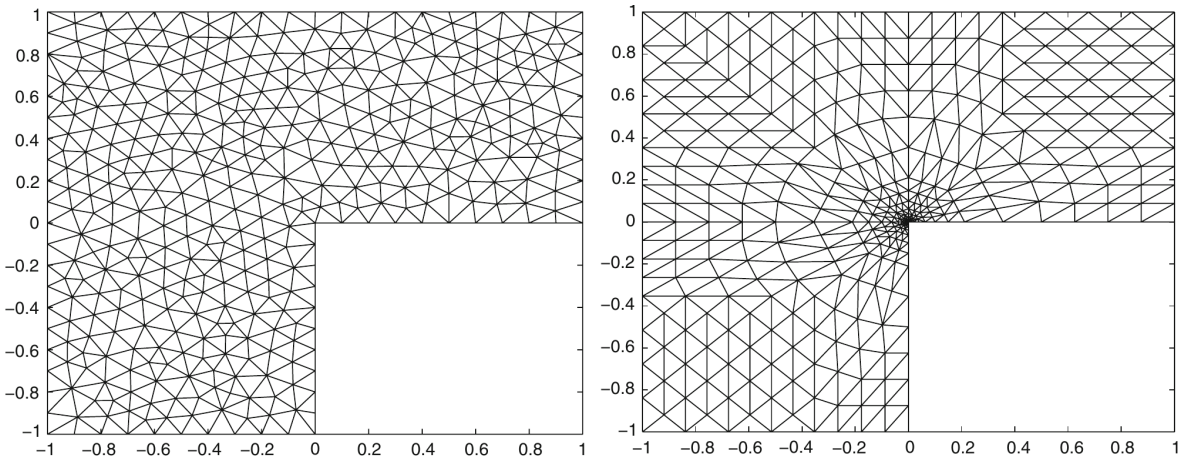


Figure 5.1: Uniform mesh (left) and graded mesh (right).

h	$r_{1,h}$	$r_{2,h}$	$r_{3,h}$	$r_{4,h}$	$r_{5,h}$
Graded 1	2.3×10^{-3}	7.9×10^{-5}	2.1×10^{-4}	2.5×10^{-4}	1.9×10^{-4}
Graded 2	5.4×10^{-4}	1.8×10^{-5}	4.7×10^{-1}	5.1×10^{-5}	8.5×10^{-2}
Graded 3	1.6×10^{-4}	4.5×10^{-6}	2.0×10^{-2}	1.3×10^{-5}	4.6×10^{-2}

Table 5.2: Relative errors on graded meshes with $\alpha = 0.95$.

5.6 Propagation in a waveguide [A3]

We want to solve problem (5.1.19)–(5.1.24), where the field $\mathbf{E} \in \mathbf{X}_{N,\Gamma_C}(\Omega)$. According to Costabel [112], the graph norm and the semi-norm are equivalent on $\mathbf{X}_{N,\Gamma_C}(\Omega)$, defined in (5.1.25):

$$\|\mathbf{v}\|_{\mathbf{X}_{N,\Gamma_C}(\Omega)} = \left(\|\operatorname{\mathbf{curl}} \mathbf{v}\|^2_{\mathbf{L}^2(\Omega)} + \|\operatorname{div} \mathbf{v}\|^2_{L^2(\Omega)} + \|\mathbf{v} \times \mathbf{n}\|^2_{\mathbf{L}^2(\Gamma_A)} \right)^{\frac{1}{2}}.$$

5.6.1 Variational formulations

Starting from the second-order in time system of Equations (5.1.19)-(5.1.24), we obtain a series of equivalent variational formulations. When Γ_A is empty, we refer the interested reader to [B2, chap. 12] for details. When Γ_A is non-empty, it is enough to adapt the proofs given by Ben Belgacem and Bernardi [32] to handle the boundary terms. The three variational formulations are obtained successively this way:

1° Multiply Equation (5.1.19) by $\mathbf{v} \in \mathbf{H}_{0,\Gamma_C}(\mathbf{curl}, \Omega)$, and integrate by parts over Ω. We get the variational formulation:
Find $\mathbf{E}(t) \in \mathbf{H}_{0,\Gamma_C}(\mathbf{curl}, \Omega)$ such that for all $\mathbf{v} \in \mathbf{H}_{0,\Gamma_C}(\mathbf{curl}, \Omega)$, for all t,

$$\langle \partial^2_{tt}\mathbf{E}, \mathbf{v}\rangle + c^2(\operatorname{\mathbf{curl}} \mathbf{E}, \operatorname{\mathbf{curl}} \mathbf{v})_{\mathbf{L}^2(\Omega)} + c\,(\partial_t \mathbf{E} \times \mathbf{n}, \mathbf{v} \times \mathbf{n})_{\mathbf{L}^2(\Gamma_A)} = -(\partial_t \mathbf{J}/\varepsilon_0, \mathbf{v})_{\mathbf{L}^2(\Omega)} + c\,(\partial_t \mathbf{b} \times \mathbf{n}, \mathbf{v})_{\mathbf{L}^2(\Gamma_A)}.$$

Above, $\langle \cdot, \cdot \rangle$ denotes an ad hoc duality product between $\partial^2_{tt}\mathbf{E}$ and elements of $\mathbf{H}_{0,\Gamma_C}(\mathbf{curl}, \Omega)$, cf. [32]. We have the

Theorem 5.6.1. *Suppose that $\partial_t \mathbf{J} \in L^2(0,T;\mathbf{L}^2(\Omega))$, $\varrho \in \mathcal{C}^0(0,T;H^{-1}(\Omega))$, ϱ and $\mathbf{J}$ satisfying* (5.1.5). *Suppose that $(\mathbf{E}_0, \mathbf{E}_1) \in \mathbf{H}_{0,\Gamma_C}(\mathbf{curl}, \Omega) \times \mathbf{L}^2(\Omega)$. Then, Equation* (5.6.1)*, together with initial conditions* (5.1.23)-(5.1.24)*, is equivalent to problem* (5.1.19)-(5.1.24) *and has a unique solution $\mathbf{E}$ such that $(\mathbf{E}, \partial_t \mathbf{E}) \in \mathcal{C}^0(0,T;\mathbf{H}_{0,\Gamma_C}(\mathbf{curl}, \Omega)) \times \mathcal{C}^0(0,T;\mathbf{L}^2(\Omega))$. Moreover, provided that $\varrho \in \mathcal{C}^0(0,T;L^2_\alpha(\Omega))$ and $\mathbf{E}_0 \in \mathbf{X}^\alpha_{N,\Gamma_C}(\Omega)$, we have the improved regularity result $\mathbf{E} \in \mathcal{C}^0(0,T;\mathbf{X}^\alpha_{N,\Gamma_C}(\Omega))$.*

2° Add $c^2(\operatorname{div} \mathbf{E}, \operatorname{div} \mathbf{v})_{L^2_\alpha(\Omega)}$ on the left hand side and $c^2(\varrho/\varepsilon_0, \operatorname{div} \mathbf{v})_{L^2_\alpha(\Omega)}$ on the right hand side to get the *augmented variational formulation*:
Find $\mathbf{E}(t) \in \mathbf{X}^\alpha_{N,\Gamma_C}(\Omega)$ such that for all $\mathbf{v} \in \mathbf{X}^\alpha_{N,\Gamma_C}(\Omega)$, for all t,

$$\begin{aligned} &(\partial^2_{tt}\mathbf{E}, \mathbf{v})_{\mathbf{L}^2(\Omega)} + c^2(\mathbf{E}, \mathbf{v})_{\mathbf{X}^\alpha_N(\Omega)} + c\,(\partial_t \mathbf{E} \times \mathbf{n}, \mathbf{v} \times \mathbf{n})_{\mathbf{L}^2(\Gamma_A)} \\ &= -(\partial_t \mathbf{J}/\varepsilon_0, \mathbf{v})_{\mathbf{L}^2(\Omega)} + c^2(\varrho/\varepsilon_0, \operatorname{div} \mathbf{v})_{L^2_\alpha(\Omega)} + c(\mathbf{b}' \times \mathbf{n}, \mathbf{v})_{\mathbf{L}^2(\Gamma_A)}. \end{aligned} \tag{5.6.1}$$

Theorem 5.6.2. *Suppose that $\partial_t \mathbf{J} \in L^2(0,T;\mathbf{L}^2(\Omega))$, $\varrho \in \mathcal{C}^0(0,T;L^2_\alpha(\Omega))$, ϱ and $\mathbf{J}$ satisfying* (5.1.5). *Suppose that $(\mathbf{E}_0, \mathbf{E}_1) \in \mathbf{X}^\alpha_{N,\Gamma_C}(\Omega) \times \mathbf{L}^2(\Omega)$. Then, Equation* (5.6.1)*, together with initial conditions* (5.1.23)–(5.1.24)*, is equivalent to problem* (5.1.19)–(5.1.24) *and has a unique solution $\mathbf{E}$ such that $(\mathbf{E}, \partial_t \mathbf{E}) \in \mathcal{C}^0(0,T;\mathbf{X}^\alpha_{N,\Gamma_C}(\Omega)) \times \mathcal{C}^0(0,T;\mathbf{L}^2(\Omega))$.*

3° Add $(p, \operatorname{div} \mathbf{v})_{L^2_\alpha(\Omega)}$ on the left hand side and consider a constraint on $\operatorname{div} \mathbf{E}$ (cf. (5.6.3)). Let $p \in L^2_\alpha(\Omega)$ be the Lagrange multiplier, we reach the *mixed augmented variational formulation*:
Find $(\mathbf{E}(t), p(t)) \in \mathbf{X}^\alpha_{N,\Gamma_C}(\Omega) \times L^2_\alpha(\Omega)$ such that for all $\mathbf{v} \in \mathbf{X}^\alpha_{N,\Gamma_C}(\Omega)$, for all t,

$$\begin{aligned} &(\partial^2_{tt}\mathbf{E}, \mathbf{v})_{\mathbf{L}^2(\Omega)} + c^2(\mathbf{E}, \mathbf{v})_{\mathbf{X}^\alpha_{N,\Gamma_C}(\Omega)} + (p, \operatorname{div} \mathbf{v})_{L^2_\alpha(\Omega)} + (\partial_t \mathbf{E} \times \mathbf{n}, \mathbf{v} \times \mathbf{n})_{\mathbf{L}^2(\Gamma_A)} \\ &\qquad = -(\mathbf{J}'/\varepsilon_0, \mathbf{v})_{\mathbf{L}^2(\Omega)} + c^2(\varrho/\varepsilon_0, \operatorname{div} \mathbf{v})_{L^2_\alpha(\Omega)} + \int_{\Gamma_A} (c\mathbf{b}' \times \mathbf{n}).\mathbf{v}, \end{aligned} \tag{5.6.2}$$

and for all $q \in L^2_\alpha(\Omega)$, for all t,

$$(\operatorname{div} \mathbf{E}, q)_{L^2_\alpha(\Omega)} = (\varrho/\varepsilon_0, q)_{L^2_\alpha(\Omega)}. \tag{5.6.3}$$

Theorem 5.6.3. *Suppose that $\partial_t \mathbf{J} \in L^2(0,T;\mathbf{L}^2(\Omega))$, $\partial_t \varrho \in L^2(0,T;L^2_\alpha(\Omega))$, ϱ and $\mathbf{J}$ satisfying* (5.1.5). *Suppose that $(\mathbf{E}_0, \mathbf{E}_1) \in \mathbf{X}^\alpha_{N,\Gamma_C}(\Omega) \times \mathbf{L}^2(\Omega)$. Then, Equations* (5.6.2)-(5.6.3) *are equivalent to Equa-*

tions (5.1.19)-(5.1.24) *and have a unique solution* $(\mathbf{E}, p)$ *such that* $(\mathbf{E}, \partial_t \mathbf{E}) \in \mathcal{C}^0(0, T; \mathbf{X}^\alpha_{N,\Gamma_C}(\Omega)) \times \mathcal{C}^0(0, T; \mathbf{L}^2(\Omega))$ *and* $p = 0$.

The constraint (5.6.3) is added to reinforce Gauss' law (5.1.20) and also to avoid numerical instabilities when the discrete charge conservation equation is not satisfied while solving the Maxwell-Vlasov system. Actually, if the charge conservation does not hold, then the Lagrange multiplier no longer vanishes, and it compensates for the discrepancy [27]. One can build similar variational formulation for the magnetic fields.

5.6.2 Discretization

Let us discretize Problem (5.6.2)–(5.6.3). Concerning the time discretization, we use explicit centered schemes of order two. Let Δt be the time step and $t_n = n\Delta t$, $n \in \{0, \dots, T/\Delta t\}$ be the discrete times, we consider that:

- $\partial_t u(\cdot, t_n)$ is approximated by $[u(\cdot, t_{n+1}) - u(\cdot, t_{n-1})]/(2\Delta t)]$.
- $\partial^2_{tt} u(\cdot, t_n)$ is approximated by $[u(\cdot, t_{n+1}) - 2u(\cdot, t_n) + u(\cdot, t_{n-1})]/(\Delta t)^2]$

Concerning the space discretization, we use the $\mathbf{P}^k$ finite elements for (5.6.1) ($k \geq 1$) and the $\mathbf{P}^k \times P^{k-1}$ Taylor-Hood Finite Element [239] ($k \geq 2$). We denote by $(\mathbf{E}^n_h, p^n_h)$ the discretization of $(\mathbf{E}(t_n), p(t_n))$. We recall that for the fully discretized explicit scheme, on must satisfy a CFL-like condition. For $\mathbf{P}^k$ finite elements, we must have: $c\,\Delta t \leq C_k \min_\ell \rho_\ell$, where C_k depends on the approximation order and ρ_ℓ is defined in §4.7. With this discretization, the $\mathbf{L}^2(\Omega)$ error estimate reads [214, chap. 8]:

$$\max_n(\|\mathbf{E}(t_n) - \mathbf{E}^n_h\|_{\mathbf{L}^2(\Omega)}) \leq C_\mu \,(\,(\Delta t)^2 + h^\mu), \quad \forall \mu < \tau_\alpha.$$

Additionally, we note that if one considers an implicit scheme such as the Crank-Nicholson scheme, one can obtain, following the result [105]:

$$\max_n(\|\partial_t \mathbf{E}(t_n) - \partial \mathbf{E}^n_h\|^2_{\mathbf{L}^2(\Omega)} + c^2 \|\mathbf{E}(t_n) - \mathbf{E}^n_h\|_{\mathbf{X}^\alpha_N}) \leq C_\mu \,(\,(\Delta t)^2 + h^{2\mu} + (\Delta t)^2\, h^{2(\mu-1)}), \quad \forall \mu \leq \tau_\alpha,$$

where $\partial_t \mathbf{E}^n_h = (\mathbf{E}^n_h - \mathbf{E}^{n-1}_h)/\Delta t$.

5.6.3 Numerical scheme

Let $N_\mathbf{E}$ (resp. N_p) be the number of $\mathbf{P}^{k+1}$ (resp. P^k) degrees of freedom for the electric field (resp. the Lagrange multiplier). Let $\underline{E}^n \in \mathbb{R}^{N_\mathbf{E}}$ (resp. $\underline{p}^n \in \mathbb{R}^{N_p}$) be the discretized electric field (resp. Lagrange multiplier) at time t_n. Let $\mathbb{M}_\Omega \in \mathbb{R}^{N_\mathbf{E} \times N_\mathbf{E}}$ be the mass matrix and $\mathbb{M}^{\|}_\Gamma \in \mathbb{R}^{N_\mathbf{E} \times N_\mathbf{E}}$ be the boundary mass matrix, defined on Γ_A. We define $\mathbb{M} = \mathbb{M}_\Omega + \frac{1}{2} c\,\Delta t\, \mathbb{M}^{\|}_{\Gamma_A}$. Let $\mathbb{B} \in \mathbb{R}^{N_p \times N_\mathbf{E}}$ be the constraint matrix. At time t_n, we have to solve:
Find $(\mathbf{E}^{n+1}, p^{n+1}) \in \mathbb{R}^{N_\mathbf{E}} \times \mathbb{R}^{N_p}$ such that

$$\mathbb{M}\underline{E}^{n+1} + (\Delta t)^2 \mathbb{B}^T \underline{p}^{n+1} = \underline{F}^{n+1/2}, \text{ and } \mathbb{B}\underline{E}^{n+1} = \underline{G}^{n+1}. \tag{5.6.4}$$

The algorithm is the same as algorithm (4.5.8) with $\mathbb{A}_u = \mathbb{M}$, $\mathbb{B} = \mathbb{B}_{p,u}$, $\mathbb{K}_p = \mathbb{B}\mathbb{M}^{-1}\mathbb{B}^T$ and $\underline{U} = \underline{E}^{n+1}$, $\underline{F} = \underline{F}^{n+1/2}$, $\underline{G} = \underline{G}^{n+1}$:

$$\begin{array}{ll} \text{Prediction:} & \text{Solve in } \underline{E}_\star \text{ such that } \mathbb{M}\,\underline{E}_\star = \underline{F}^{n+1/2}. \\ \text{Constraint solver:} & \text{Solve in } \underline{P} \text{ such that } \mathbb{K}_p\,\underline{P} = \mathbb{B}\,\underline{E}_\star - \underline{G}^{n+1}. \\ \text{Correction:} & \text{Solve in } \delta\underline{E} \text{ such that } \mathbb{M}\,\delta\underline{E} = -\mathbb{B}^T\,\underline{P}^{n+1}. \\ \text{Update:} & \underline{E}^{n+1} = \delta\underline{E} + \underline{E}_\star. \end{array} \tag{5.6.5}$$

One can use the P^k mass matrix as a preconditioner to solve the second step with the Uzawa algorithm (see §4.5.3). When the domain is convex, this yields an optimal method in term of number of iterations, which does not depend on the meshsize as proved in [B2]. When the domain is not convex, it has been observed numerically that this number of iterations grows, albeit very slowly, when h decreases. Moreover, in practice, given a mesh made of tetrahedra with good aspect ratio, it has been noted [157] that $\underline{p}^{n+1}$ doesn't vary much (from the zero theoretical value), so that there is actually no need to compute it at all times. However, if there is a coupling with the Vlasov equation, one has to compute accurately [27] this Lagrange multiplier. To speed up the resolution further, one can perform the so-called mass-lumping techniques, to obtain diagonal mass matrices in (5.6.4). It has been established in [111] that this could be achieved, for a continuous discretization, with no loss in accuracy. This results in the so-called $\tilde{P}^1$ or $\tilde{P}^2$ finite elements. In the latter case, accuracy is preserved at the cost of increasing the total number of degrees of freedom, going from 10 degrees of freedom for the P^2 finite element to 23 degrees of freedom for the $\tilde{P}^2$ finite element. In the former case, the number of dof is equal to four for both the P^1 and the $\tilde{P}^2$ finite elements.

5.6.4 Numerical illustration in a non-convex geometry

The domain Ω is an extruded, L-shaped, polyhedron (see Figure 5.2), of depth $L = 4\,\mathrm{m}$. Geometrically speaking, its boundary $\partial\Omega$ has a single reentrant edge of dihedral angle $\Theta_e = 2\pi/3$ (in red). In this case, it can be checked that $\alpha_{\min} = 1 - \Theta_e/\pi = 1/3$. Physically speaking, the boundary is split in two parts, Γ_C and Γ_A (in gray). The latter is composed of the leftmost and rightmost faces of the boundary. Further, the Silver–Müller boundary condition (2.8) on the artificial boundary Γ_A holds with $\mathbf{b} = 0$: we impose an absorbing condition. The data are made of zero initial conditions $(\mathbf{E}_0, \mathbf{B}_0)$, and non-zero charge and current densities $\varrho, \mathbf{J}$). A current bar B_J crosses the domain, and inside this bar, there holds (with $\omega = 2.5\,\mathrm{GHz}$):

$$\mathbf{J} = 10^{-5}\omega\,\sin(\pi z/L)\cos(\omega\,t)\mathbf{e}_z, \quad \varrho = 10^{-5}(\pi/L)\,\cos(\pi z/L)\cos(\omega\,t).$$

The wavelength and time period associated to ω are, respectively, equal to $2\pi c/\omega \approx 0.75\,\mathrm{m}$ and $2\pi/\omega \approx 2.5\,\mathrm{ns}$. It is clear that the dimensions of our domain are not realistic; however, we made this choice in order to be able to visualize several oscillations of the field (in space). We note that both $\mathbf{J}$ and ϱ are smooth, so all theorems of §5.6.1 apply, which ensures theoretically the existence and uniqueness of the electromagnetic field.

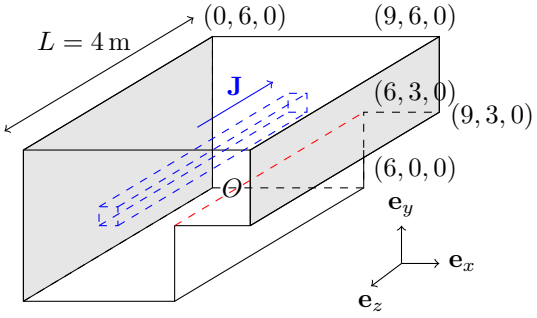


Figure 5.2: The model problem.

To compute the electric field, we considered the space $\mathbf{X}^{\alpha}_{N,\Gamma_C}(\Omega)$, with $\alpha = 0.95$ and the augmented variational formulation (5.6.1). The value of the exponent $\alpha - \alpha_{\min}$ that appears in the convergence rates is then approximately equal to 0.62. To compute the magnetic induction, we introduced (again with $\alpha = 0.95$):

$$\mathbf{X}^{\alpha}_{T,\Gamma_C}(\Omega) := \{\mathbf{v} \in \mathbf{H}(\mathbf{curl},\Omega)\,|\,\mathrm{div}\,\mathbf{v} \in L^2_\alpha(\Omega),\ \mathbf{v}\times\mathbf{n}_{|\partial\Omega} \in \mathbf{L}^2_t(\partial\Omega),\ \mathbf{v}\cdot\mathbf{n}_{|\Gamma_C} = 0\}.$$

The augmented variational formulation reads:
Find $\mathbf{B}(t) \in \mathbf{X}^{\alpha}_{T,\Gamma_C}(\Omega)$ such that for all $\mathbf{v} \in \mathbf{X}^{\alpha}_{T,\Gamma_C}(\Omega)$, for all t,

$$\langle \partial^2_{tt}\mathbf{B}, \mathbf{v}\rangle + c^2(\mathbf{B},\mathbf{v})_{\mathbf{X}^{\alpha}_{T,\Gamma_C}(\Omega)} + c(\mathbf{B}\times\mathbf{n}, \mathbf{v}\times\mathbf{n})_{\mathbf{L}^2(\Gamma_A)} = (1/\varepsilon_0)\,(\mathbf{J}, \mathbf{curl}\,\mathbf{v})_{\mathbf{X}^{\alpha}_{T,\Gamma_C}(\Omega)}.$$

We report the results of computations made with the $\tilde{P}^1$ finite elements on a tetrahedral mesh, with 685×10^3 tetrahedra. In particular, the 10 discretization point per wavelength rule is satisfied. One has $\min_\ell \rho_\ell \approx 6\,\mathrm{cm}$, and the time-step is in the order of $\Delta t \approx 40\,\mathrm{ps}$: this complies with the CFL condition. The final observation time is $T = 25\,\mathrm{ns}$: it is sufficient for the electromagnetic wave that is generated by B_J to span the whole domain (recall that the wave travels at the finite speed $c \approx 3 \times 10^8\,\mathrm{m\,.\,s^{-1}}$.

In Figure 5.3, we present the time evolution of the x-component of the electric field at the locations $M_1 = (1,1,2)$, $M_2 = (1,5,2)$, $M_3 = (5.5, 2.5, 2)$, $M_4 = (8, 5.5, 2)$; respectively. It remains equal to zero until the electric wave reaches the point under consideration. Then the field oscillates with a period $\approx 2.5\,\mathrm{ns}$: as expected, we observe forced oscillations.

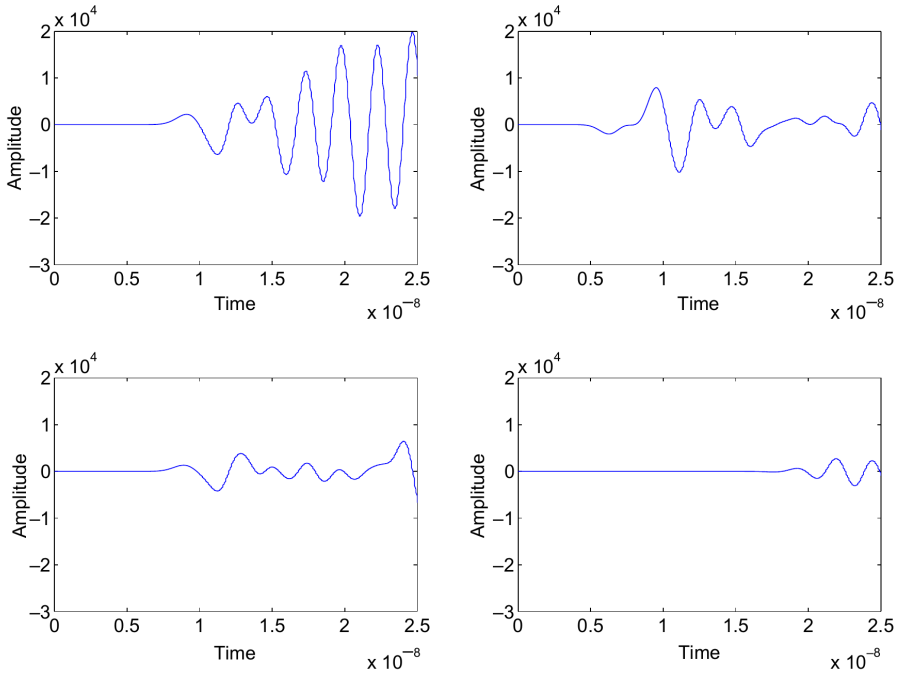


Figure 5.3: $\mathbf{E}_x$ at $(M_i)_{i=1}^4$.

Let us now focus on the spatial behaviour of the electromagnetic field, which we expect to be singular[5], in the neighbourhood of the reentrant edge. We report the evolution of the field in a plane, which is perpendicular to that edge: the geometrical singularity projects to a reentrant corner.

In Figure 5.4 and 5.5, the space evolution of the x and y-components of the electric field is represented in the plane $z = 2.5\,\mathrm{m}$, at times $T_1 = 1\,\mathrm{ns}$, $T_2 = 8\,\mathrm{ns}$, $T_3 = 15\,\mathrm{ns}$, $T_4 = 20\,\mathrm{ns}$. We can see that an electric wave is created by the current, that it propagates into the cavity with wavelength $\approx 0.75\,\mathrm{m}$, and is reflected by the conductor as expected. At T_3, we observe a growing peak of intensity close to the reentrant corner. We also provide a close up of the x-component near the reentrant corner in Figure 5.6, again at times $(T_i)_{i=1}^4$. The usual pattern of strong variations is visible, once the wave has reached the corner, especially at time T_3.

In Figure 5.7, we picture the space evolution of the z-component in the plane $z \approx 2.5\,\mathrm{m}$, at times $(T_i)_{i=1}^4$. Again, we observe the propagation of the wave with wavelength $\approx 0.75\,\mathrm{m}$, and the reflections.

[5] According to physics principles, to experiments, and also to the mathematical theory.

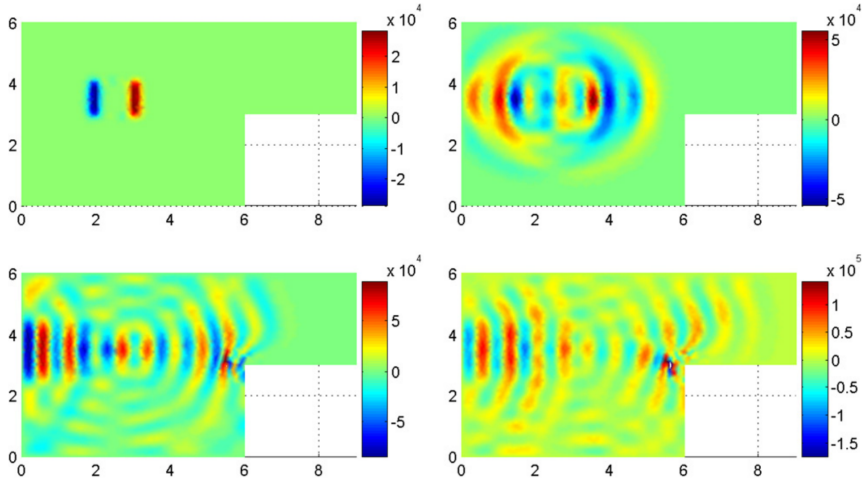


Figure 5.4: $\mathbf{E}_x$ at time $(T_i)_{i=1}^4$ in the plane $z = 2.5\,\mathrm{m}$.

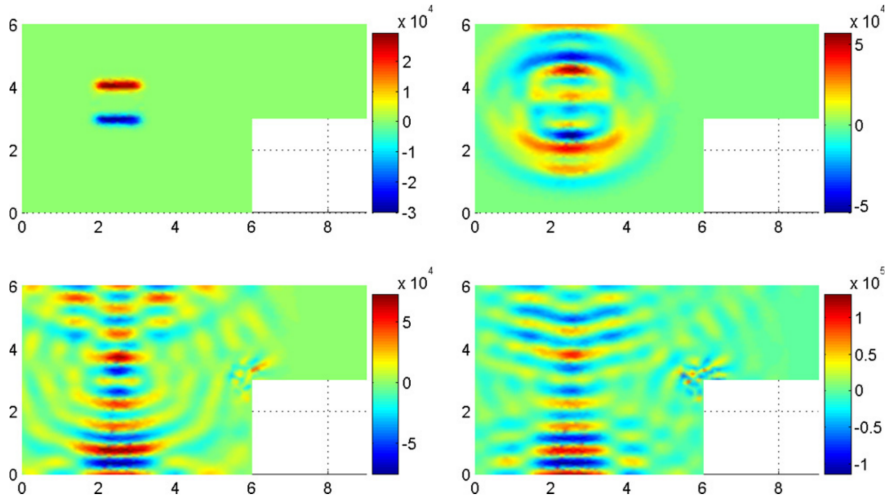


Figure 5.5: $\mathbf{E}_y$ at time $(T_i)_{i=1}^4$ in the plane $z = 2.5\,\mathrm{m}$.

Note that this component is always regular (as opposed to singular), which is due to the fact that the geometrical singularity is parallel the z axis [75, 98]. Moreover, it takes smaller (absolute) values than the x- and y-components.

In Figs. 5.4, 5.5 and 5.7, one can see spurious reflections on Γ_A, due to the fact that the Silver–Müller boundary condition is simply of first order: only plane waves with normal incidence are absorbed, which is not our case. As far as the magnetic induction is concerned, we observe a similar behaviour up to a rotation of $\pi/2$ around the z-axis. So for instance, we show in Figure 5.8 the space evolution of $\mathbf{B}_y$ in the plane $z \approx 2.5\,\mathrm{m}$, at times $(T_i)_{i=1}^4$.

Physically speaking, the overall space distribution of the electromagnetic field can be explained with the help of the Biot-Savart law. According to this law, the magnetic induction created by the current bar B_J at location M reads:

$$\mathbf{B} = \frac{\mu_0}{4\pi} \int_{B_J} \frac{\mathbf{J} \times (\mathbf{x}_M - \mathbf{x})}{|\mathbf{x}_M - \mathbf{x}|^2} \text{ where } \mathbf{x}_M := \vec{OM}.$$

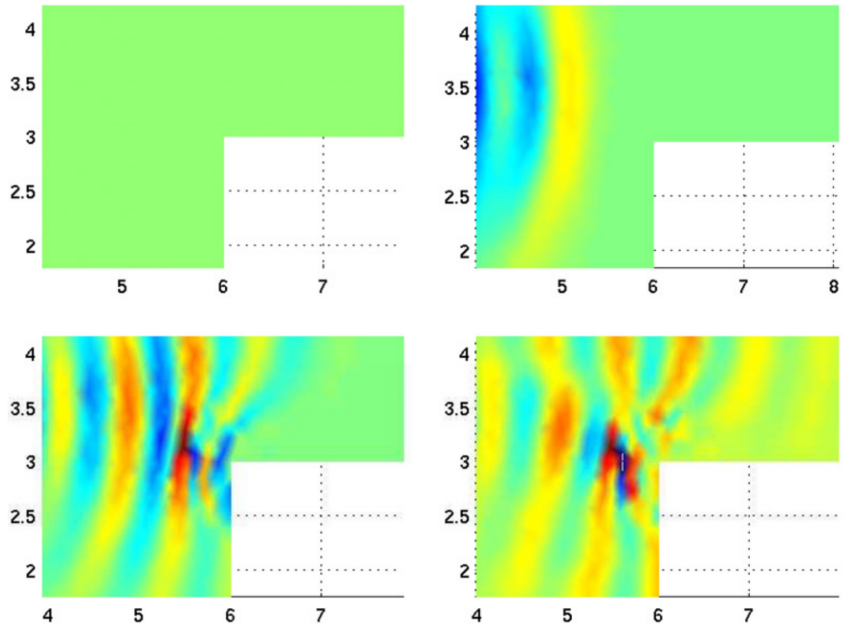


Figure 5.6: A close up of $\mathbf{E}_x$ at time $(T_i)_{i=1}^4$ in the plane $z = 2.5\,\mathrm{m}$.

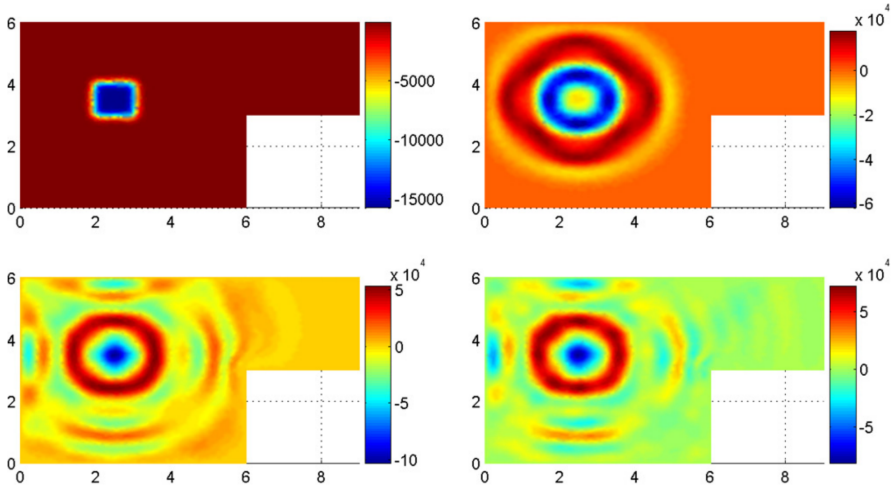


Figure 5.7: $\mathbf{E}_z$ at time $(T_i)_{i=1}^4$ in the plane $z = 2.5\,\mathrm{m}$.

Let (x, y, z) be the coordinates of M, and let (x, y, z) be those of M', the orthogonal projection of M on the z-axis of B_j. One gets:

$$\mathbf{B}(\mathbf{x}_M) \propto \frac{J_z}{|\mathbf{x}_M - \mathbf{x}_{M'}|^2} \begin{pmatrix} -(y' - y) \\ (x - x') \end{pmatrix}$$

Therefore, if $y = y'$, $B_x(\mathbf{x}_M) \approx 0$, whereas if $x = x'$, $B_y(\mathbf{x}_M) \approx 0$. In Figure 5.8, one sees in particular that, at time T_1, B_y vanishes vertically, i.e. for those locations M such that $x = x0'$, that is above and below the current bar B_J. The electric field being orthogonal to the magnetic induction, one gets that if $y = y'$, $E_y(\mathbf{x}_M) \approx 0$, whereas if $x = x'$, $E_x(\mathbf{x}_M) \approx 0$. This can be observed in Figs. 5.4 and 5.5.

Numerically speaking, we note that if one implements the augmented variational formulations with the plain scalar product $(\cdot,\cdot)_{\mathbf{X}_N(\Omega)}$ instead of the weighted $(\cdot,\cdot)_{\mathbf{X}_N^\alpha(\Omega)}$, one gets similar numerical results until the wave hits the reentrant edge. This is a consequence of the fact that before that time, the

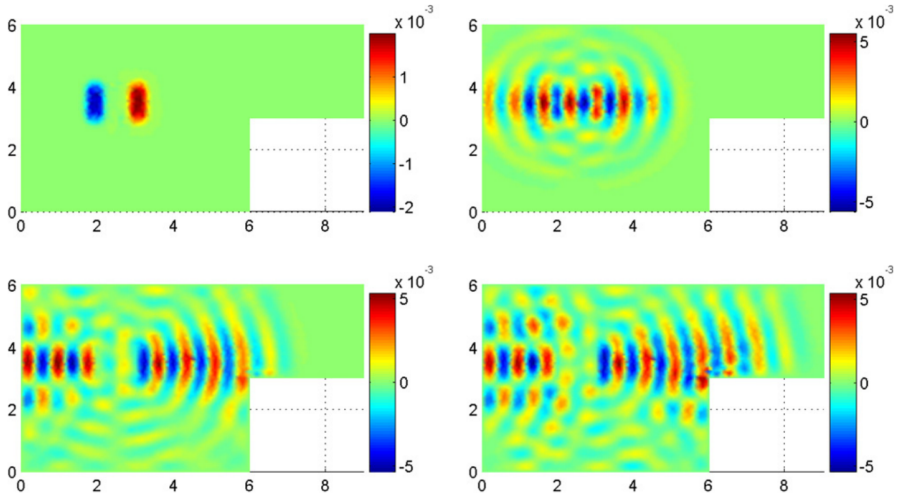


Figure 5.8: $\mathbf{B}_y$ at time $(T_i)_{i=1}^4$ in the plane $z = 2.5 \approx$ m.

fields are H^1-regular: both augmented variational formulations have the same solution. After that, the fields become singular, and results differ considerably: first, in the neighbourhood of this edge, and this discrepancy extends in space at the speed of light c. This is due to the fact that the augmented variational formulation with the plain scalar product $(\cdot,\cdot)_{\mathbf{X}_N(\Omega)}$ cannot capture the singular part of the fields. In other words, it is crucial that one uses the correct variational formulation, in the correct functional space.

Chapter 6

Neutronics

The work presented in this Chapter concern articles [A12, A2, A6, A7, A11, A10], proceedings [P5, P6], Chapter 3 of lecture notes [B1] and PhD thesis of Léandre Giret [163]. These references are recalled below, in reverse chronological order:

[A12] *Numerical analysis of the neutron multigroup SP_N equations*,
E. Jamelot and F. Madiot, Comptes Rendus Mathématique, **359**, 5, pp. 533–545, 2021.

[A2] *Numerical analysis of the mixed finite element method for the neutron diffusion problem with highly heterogeneous coefficients*, P. Ciarlet Jr., L. Giret, E. Jamelot, F. Kpadonou, ESAIM: Mathematical Modelling and Numerical Analysis, **52**, 5, pp. 2003–2035, 2018.

[A6] *Domain decomposition methods for the diffusion equation with low-regularity solution*,
P. Ciarlet Jr., E. Jamelot and F. D. Kpadonou,
Computers & Mathematics with Applications, **74**, 10, pp. 2369–2384, 2017.

[A7] *A simple monodimensional model coupling an enthalpy production equation and a neutronic diffusion equation*, S. Dellacherie, E. Jamelot, O. Lafitte,
Applied Mathematics Letters, **62**, pp. 35–41, 2016.

[A11] *Fast non-overlapping Schwarz domain decomposition methods for solving the neutron diffusion equation*, E. Jamelot and P. Ciarlet Jr.,
Journal of Computational Physics, **241**, pp. 445–463, 2013.

[A10] *Domain Decomposition for the SP_N Solver MINOS*,E. Jamelot, A.-M. Baudron, J.-J. Lautard,
Transport Theory and Statistical Physics, **41** pp. 495–512, 2012.

[P5] *Critical Computation with Finite Element Method on Non-Conforming Meshes*,
L. Giret, P. Ciarlet Jr., E. Jamelot, M&C 2017, April 16-20, Jeju (Korea).

[P2] *Numerical results for the coupling of a simple neutronics diffusion model and a simple hydrodynamics low mach number model without coupling codes*,
S. Dellacherie, E. Jamelot, O. Lafitte, R. Mouhamad,
SYNASC 2016, September 24-27, Timişoara (Romania).

[P6] *Domain decomposition for the neutron SP_N equations*,
E. Jamelot, P. Ciarlet Jr., A.-M. Baudron, J.-J. Lautard,
DDM'21, 2012, June 25-29, Rennes (France).

[B1] *Outils mathématiques et algorithmiques pour le calcul scientifique*, P. Ciarlet Jr., E. Jamelot, (pp. 185-229), ENSTA Paris, 2019.

[163] *Numerical Analysis of a Non-Conforming Domain Decomposition for the Multigroup SP_N Equations*, L. Giret, PhD thesis, Université Paris-Saclay, 2018.

Article [A6] followed the Master's 2 intership of Félix D. Kpadonou. Article [A2] and proceedings [P5] are related to the PhD thesis of Léandre Giret [163]. Sofiane Ezzehi's internship provided an opportunity to expand on the coupling strategy described in article [A7] and proceedings [P2].

This Chapter is organized as follows: in Section 6.1, we give details on how to obtain the neutron simplified transport equations. Next, in Section 6.2, we carry out the numerical analysis of the neutron diffusion source problem discretized with Raviart-Thomas finite elements. This is followed, in Section 6.3, by the numerical analysis of the neutron diffusion generalized eigenproblem. In Section 6.4, we give tools to apply domain decomposition methods. We first study the optimized Schwarz method in Section 6.5, and then the so-called DD+L^2 jumps method in Section 6.6. In Section 6.7, we carry out the numerical analysis of the neutron multigroup simplified transport equation, discretized with discontinuous Galerkin finite elements. Finally, in Section 6.8, we study a simplified thermal-hydraulics and neutronics coupling.

6.1 Context and model

A nuclear reactor produces thermal energy that is released from induced nuclear fission in fissile atoms, such as uranium. Fission produces kinetic energy, γ rays, lighter atomic nuclei and free neutrons. These neutrons can lead to further fission: this process is known as nuclear fission chain reaction. To calculate the power of a nuclear reactor, we study the stationary state of its core, which is where fission occurs. This reaction is modelled by the neutron transport equation, which represents the neutron balance in the reactor.

6.1.1 Neutron transport equation

The neutron density, $\mathcal{N}(t, \mathbf{x}, \omega, E)$, depends on seven variables: time $t \in (0, T)$, position in space $\mathbf{x} \in \mathbb{R}^3$, velocity direction $\mathbf{\Omega} \in \mathbb{S}^2$ ($\mathbb{S}^2$ is the unit sphere), and energy $E \in \mathbb{R}^+$. The energy of a neutron is related to its velocity v by the kinetic energy equation: $E = \frac{1}{2} m\, v^2$, m being the mass of the neutron. The data from the neutron transport equation are the macroscopic cross sections, as well as the fission yield $\nu(E)$ and the fission spectrum $\chi(E)$. The macroscopic scattering cross section is denoted $\sigma_s(t, \mathbf{x}, \widehat{(\mathbf{\Omega}, \mathbf{\Omega}')}, E' \to E)$. It represents the probability that a neutron colliding at time t and point $\mathbf{x}$, with energy E' and direction $\mathbf{\Omega}'$, will be reflected with energy E and direction $\mathbf{\Omega}$. Shocks are assumed to be isotropic, so σ_s depends on the difference between the neutron's angles of incidence $\mathbf{\Omega}'$ and reflection $\mathbf{\Omega}$. The macroscopic fission cross section, denoted σ_f, represents the probability that a neutron colliding at time t and with energy E' at the point $\mathbf{x}$ will cause fission. This event does not depend on the incident angle and is isotropic. The total macroscopic cross-section σ_t represents the probability of a neutron of energy E undergoing a collision at time t and point $\mathbf{x}$. The macroscopic scattering cross section models neutron collisions. It depends on the angle between the angles of incidence $\mathbf{\Omega}'$ and reflection $\mathbf{\Omega}$ of the neutron, as well as the energy before E' and after E the shock. The angular neutron flux is given by the quantity $\psi = v\,\mathcal{N}$, and the angular neutron current by the vector $\mathcal{J} = \mathbf{\Omega}\,\psi$.

The neutron transport equation reads:

$$\begin{aligned} \frac{\partial \mathcal{N}}{\partial t} \;=\; & \int_0^{+\infty} \int_{\mathbb{S}^2} \sigma_s \left(t, \mathbf{x}, \widehat{(\mathbf{\Omega}, \mathbf{\Omega}')}, E' \to E\right) \psi(t, \mathbf{x}, \mathbf{\Omega}', E') \mathrm{d}\mathbf{\Omega}' \mathrm{d}E' \\ & + \frac{\chi(E)}{4\pi} \left(\int_0^{+\infty} \nu(E')\, \sigma_f(t, \mathbf{x}, E') \int_{\mathbb{S}^2} \psi(t, \mathbf{x}, \mathbf{\Omega}', E') \mathrm{d}\mathbf{\Omega}' \mathrm{d}E' \right) \\ & + S_{ext} - \operatorname{div} \mathcal{J}(t, \mathbf{x}, \mathbf{\Omega}, E) - \sigma_t(t, \mathbf{x}, E) \psi(t, \mathbf{x}, \mathbf{\Omega}, E). \end{aligned} \tag{6.1.1}$$

In the steady-state case, where there is no external source $S_{ext} = 0$, Eq. (6.1.1) reads:

$$\begin{aligned}
&\int_0^{+\infty} \int_{\mathbb{S}^2} \sigma_s \left(\mathbf{x}, (\widehat{\mathbf{\Omega}, \mathbf{\Omega}'}), E' \to E\right) \psi(\mathbf{x}, \mathbf{\Omega}', E') \mathrm{d}\mathbf{\Omega}' \mathrm{d}E' \\
&+ \frac{\chi(E)}{4\pi} \left(\int_0^{+\infty} \nu(E')\, \sigma_f(\mathbf{x}, E') \int_{\mathbb{S}^2} \psi(\mathbf{x}, \mathbf{\Omega}', E') \mathrm{d}\mathbf{\Omega}' \mathrm{d}E' \right) \\
&= \operatorname{div} \mathcal{J}(\mathbf{x}, \mathbf{\Omega}, E) + \sigma_t(\mathbf{x}, E) \psi(\mathbf{x}, \mathbf{\Omega}, E).
\end{aligned} \tag{6.1.2}$$

The macroscopic cross sections are obtained from the basic nuclear data, i.e. the microscopic cross sections. The evolution of the microscopic cross sections is modelled by the Bateman equations. In this way, the Bateman equations and the neutron transport equation are coupled. In order to perform calculations, the basic nuclear data are homogenised in space and energy. See [178] about the homogenization of linear Boltzmann equation when the optical parameters are highly heterogeneous in the energy variable. Therefore, the macroscopic cross sections are numerical data. Thus, Equation (6.1.2) may have no solution. To obtain a balanced solution, we rewrite Equation (6.1.2) in the form of a generalized eigenvalue problem. We want to determine the eigenpair (ψ, k_{eff}) such that:

$$\begin{aligned}
&\operatorname{div} \mathcal{J}(\mathbf{x}, \mathbf{\Omega}, E) + \sigma_t(\mathbf{x}, E) \psi(\mathbf{x}, \mathbf{\Omega}, E) \\
&= \int_0^{+\infty} \int_{\mathbb{S}^2} \sigma_s \left(\mathbf{x}, (\widehat{\mathbf{\Omega}, \mathbf{\Omega}'}), E' \to E\right) \psi(\mathbf{x}, \mathbf{\Omega}', E') \mathrm{d}\mathbf{\Omega}' \mathrm{d}E' \\
&+ \lambda \frac{\chi(E)}{4\pi} \left(\int_0^{+\infty} \nu(E')\, \sigma_f(\mathbf{x}, E') \int_{\mathbb{S}^2} \psi(\mathbf{x}, \mathbf{\Omega}', E') \mathrm{d}\mathbf{\Omega}' \mathrm{d}E' \right), \\
&k_{eff} = \frac{1}{\min_\lambda \lambda}.
\end{aligned} \tag{6.1.3}$$

We can apply the Krein-Rutman Theorem to show that the physical solution, which is necessarily positive, is the eigenfunction associated with the smallest eigenvalue which in addition is simple [183, 126]. The inverse of this smallest eigenvalue is denoted k_{eff} and called the effective multiplicative coefficient, characterises the physical state of the reactor:

- If $k_{eff} = 1$: the core reactor is in a steady state and the nuclear chain reaction is self-sustaining. The reactor is said to be critical.
- If $k_{eff} > 1$: there are more neutrons which are produced than neutrons which disappear. The chain reaction races. The reactor is said to be supercritical.
- If $k_{eff} < 1$: there are less neutrons which are produced than neutrons which disappear. The chain reaction vanishes. The reactor is said to be subcritical.

6.1.2 Energy discretization

The energy variable is discretised using the multigroup theory [143, 80]. This corresponds to homogenisation techniques in energy and space, preserving certain physical quantities. The energy is assumed to be bounded: $E \in [E_{min}, E_{max}]$, and its domain is divided into G subintervals, the energy groups, such that: $[E_{min}, E_{max}] = [E_G, E_{G-1}] \cup \ldots \cup [E_1, E_0]$ with: $E_G = E_{min}$, $E_0 = E_{max}$ et $\forall g < g'\ E_{g'} < E_g$. Let us set $\mathcal{I}_G := \{1, \ldots, G\}$. The total macroscopic cross section, the fission macroscopic cross section

and the efficiency are approximated by piecewise constant functions for the energy variable:

$$\forall g \in \mathcal{I}_G,\ \forall E \in [E_g, E_{g-1}] : \left\{ \begin{array}{rcl} \sigma_t(\mathbf{x}, E) & \approx & \sigma_t^g(\mathbf{x}), \\ \nu(E)\,\sigma_f(\mathbf{x}, E) & \approx & \nu^g\,\sigma_f^g(\mathbf{x}). \end{array} \right.$$

The same applies to the macroscopic effective cross-section of diffusion: for all $g, g' \in \mathcal{I}_G$, for all $E \in [E_g, E_{g-1}]$:

$$\int_{E_{g'}}^{E_{g'-1}} \sigma_s(\mathbf{x}, \widehat{(\mathbf{\Omega}, \mathbf{\Omega'})}, E' \to E)\mathrm{d}E' \approx \sigma_s^{g' \to g}(\mathbf{x}, \widehat{(\mathbf{\Omega}, \mathbf{\Omega'})}).$$

The quantities ψ_g and $\mathcal{J}_g$ such that for all $g \in \mathcal{I}_G$,

$$\psi_g(\mathbf{x}, \mathbf{\Omega}) := \int_{E_g}^{E_{g-1}} \psi(\mathbf{x}, \mathbf{\Omega}, E)\mathrm{d}E \text{ and } \mathcal{J}_g(\mathbf{x}, \mathbf{\Omega}) := \mathbf{\Omega}\,\psi_g(\mathbf{x}, \mathbf{\Omega})$$

are called the angular flux and angular current of energy group g. By writing $\displaystyle\int_0^{\infty} \cdot\,\mathrm{d}E = \sum_{g\mathcal{I}_G} \int_{E_g}^{E_{g-1}} \cdot\,\mathrm{d}E$ in (6.1.3), and integrating over $[E_g, E_{g-1}]$, we obtain the following coupled system of G equations: the tuple $(\,(\mathcal{J}_g)_g, (\psi_g)_g, k_{eff})$ satisfies

$$\begin{aligned} &\mathrm{div}\,\mathcal{J}_g(\mathbf{x}, \mathbf{\Omega}) + \sigma_t^g(\mathbf{x})\psi_g(\mathbf{x}, \mathbf{\Omega}) - \sum_{g' \in \mathcal{I}_G} \int_{\mathbb{S}^2} \sigma_s^{g' \to g}\left(\mathbf{x}, \widehat{(\mathbf{\Omega}, \mathbf{\Omega'})}\right) \psi^{g'}(\mathbf{x}, \mathbf{\Omega'})\mathrm{d}\mathbf{\Omega'} \\ &\qquad = \lambda \frac{\chi^g}{4\pi} \left(\sum_{g' \in \mathcal{I}_G} \nu^{g'}\,\sigma_f^{g'}(\mathbf{x}) \int_{\mathbb{S}^2} \psi^{g'}(\mathbf{x}, \mathbf{\Omega'})\mathrm{d}\mathbf{\Omega'} \right), \qquad (6.1.4) \\ &\qquad k_{eff} = \frac{1}{\min_\lambda \lambda}. \end{aligned}$$

6.1.3 SP_N approximation

Concerning the motion direction, the P_N transport equations are obtained by developing the neutron flux on the spherical harmonics from order 0 to order N. This approach is very time-consuming. The simplified P_N (SP_N) transport theory [159] was developed to address this issue. The two fundamental hypotheses to obtain the SP_N equations are that locally, the angular flux has a planar symmetry; and that the axis system evolves slowly. The neutron flux and the scattering cross sections are then developed on the Legendre polynomials. From a mathematical point of view, SP_N equations correspond to tensorized $1D$ P_N transport equations, so that some couplings are missing. Consequently, the SP_N equations do not converge to transport equations. Nevertheless, they are commonly used by physicists since their resolution is cheap in terms of computational cost. The order N is odd, and the number of SP_N odd (resp. even) moments is $\hat{N} := \frac{N+1}{2}$. We will denote by $\mathcal{I}_e$ (resp. $\mathcal{I}_o$) the subset of even (resp. odd) integers of the integer set $\{0, \ldots, N\}$. Finally, the (motion direction and energy) discretization of the neutron flux is such that there are $\hat{N} \times G$ even and odd moments of the neutron flux. Note that while modelling the core of a pressurized water reactor, the number of groups if such that $2 \leq G \lesssim 30$, physicists usually choose $N = 1$ or 3, more rarely $N = 5$.

We will denote by $\phi = (\,(\phi_m^g)_{m \in \mathcal{I}_e}\,)_{g \in \mathcal{I}_G} \in \mathbb{R}^{\hat{N} \times G}$ the set of functions containing, for all energy group g, the even moments of the neutron flux. Likewise, we will denote by $\mathbf{p} = \left(\,(\,(p_{x,m}^g)_{m \in \mathcal{I}_o}\,)_{g \in \mathcal{I}_G}\right)_{x=1}^{d} \in \left(\mathbb{R}^{\hat{N} \times G}\right)^d$ the set of functions containing the odd moments of the neutron flux.

The reactor core is modelled by a bounded, connected and open subset $\mathcal{R}$ of $\mathbb{R}^d$, $d = 1, 2, 3$, having a Lipschitz boundary which is piecewise regular. The coefficients are piecewise regular, so that we split

$\mathcal{R}$ into $N_{\mathcal{R}}$ open disjoint parts $(\mathcal{R}_i)_{i\in\mathcal{I}_{\mathcal{R}}}$, $\mathcal{I}=\{1,\dots,N_{\mathcal{R}}\}$, $N_{\mathcal{R}}\geq 1$, cf. §4.4. with Lipschitz, piecewise regular boundaries: $\overline{\mathcal{R}}=\cup_{i=1}^{N_{\mathcal{R}}}\overline{\mathcal{R}_i}$.

For a set of vector valued functions $\mathbf{q}=\left(\left((q^g_{x,m})_{m,g}\right)_{x=1}^d\right)\in\left(\mathbb{R}^{\hat{N}\times G}\right)^d$, we make the following abuse of notation:

$$\operatorname{div}\mathbf{q}=\left(\operatorname{div}(\,(q^g_{x,m})_{x=1}^d\,)\right)_{m,g}\text{ and }\mathbf{q}\cdot\mathbf{p}=\left(\sum_{x=1}^d q^g_{x,m}\,p^g_{x,m}\right)_{m,g}\in\mathbb{R}^{\hat{N}\times G}.$$

Let $\mathbf{q}\in\left(\mathbb{R}^{\hat{N}\times G}\right)^d$ and $\mathbb{M}\in\left(\mathbb{R}^{\hat{N}\times\hat{N}}\right)^{G\times G}$. We set $\mathbf{q}_x=(q^g_{x,m})_{m,g}$ and we use the notation $\mathbb{M}\,\mathbf{q}=(\mathbb{M}\,\mathbf{q}_x)_{x=1}^d$.

Given a source term S_f, the multigroup SP_N equations with zero-flux boundary conditions[1] read as a set of coupled diffusion equations, posed in mixed form:

$$\text{Find }(\mathbf{p},\phi)\text{ such that }\left\{\begin{array}{rcl}\mathbb{T}_o\,\mathbf{p}+\nabla(\mathbb{H}\,\phi) &=& 0\\ \mathbb{H}^T\,\operatorname{div}\mathbf{p}+\mathbb{T}_e\,\phi &=& S_f\end{array}\right.. \tag{6.1.5}$$

When S_f depends on ϕ, the steady state multigroup SP_N equations read as the following generalized eigenproblem:

$$\text{Find }(\lambda,\mathbf{p},\phi)\text{ such that }\left\{\begin{array}{rcl}\mathbb{T}_o\,\mathbf{p}+\nabla(\mathbb{H}\,\phi) &=& 0\\ \mathbb{H}^T\,\operatorname{div}\mathbf{p}+\mathbb{T}_e\,\phi &=& \lambda^{-1}\mathbb{M}_f\phi\end{array}\right.. \tag{6.1.6}$$

We refer to [A10, A11, P6, A6, A2, 163] for more details on the multigroup SP_N and diffusion neutron equations and the general algorithm to solve them.

The matrices $\mathbb{H}$, $\mathbb{T}_e$, $\mathbb{T}_o$, $\mathbb{M}_f\in\left(\mathbb{R}^{\hat{N}\times\hat{N}}\right)^{G\times G}$ are such that $\forall(g,g')\in\mathcal{I}_G\times\mathcal{I}_G$ ($\delta_{\cdot,\cdot}$ is the Kronecker symbol):

- $(\mathbb{H})_{g,g'}=\delta_{g,g'}\hat{H}\in\mathbb{R}^{\hat{N}\times\hat{N}}$, with $\forall(i,j)\in\{1,\dots,\hat{N}\}^2$, $\hat{H}_{i,j}=\delta_{i,j}+\delta_{i,j-1}$.
- $(\mathbb{T}_e)_{g,g}:=\mathbb{T}_e^g\in\mathbb{R}^{\hat{N}\times\hat{N}}$ denotes the even removal matrix, such that:

  $$\mathbb{T}_e^g=diag\,\left(t_0\sigma^g_{r,0},t_2\sigma^g_{r,2},\dots\right),$$

  $(\mathbb{T}_o)_{g,g}:=\mathbb{T}_o^g\in\mathbb{R}^{\hat{N}\times\hat{N}}$ denotes the odd removal matrix, such that:

  $$\mathbb{T}_o^g=diag\,\left(t_1\sigma^g_{r,1},t_3\sigma^g_{r,3},\dots\right),$$

  where $\forall m\in\mathcal{I}_{e,o}$, $\sigma^g_{r,m}:=\sigma^g_t-\sigma^{g\to g}_{s,m}$, and $\forall m>0$, $t_m>0$.
  The coefficient σ^g_t is the macroscopic total cross section of energy group g, and the coefficients $\sigma^{g\to g}_{s,m}$ denote the P_N moments of the macroscopic self scattering cross sections from energy group g to itself.
- For $g'\neq g$:
  $(\mathbb{T}_e)_{g,g'}:=-\mathbb{S}_e^{g'\to g}\in\mathbb{R}^{\hat{N}\times\hat{N}}$ denotes the even scattering matrix, such that:

  $$\mathbb{S}_e^{g'\to g}=diag\,\left(t_0\sigma^{g'\to g}_{s,0},t_2\sigma^{g'\to g}_{s,2},\dots\right),$$

  $(\mathbb{T}_o)_{g,g'}:=-\mathbb{S}_o^{g'\to g}\in\mathbb{R}^{\hat{N}\times\hat{N}}$ denotes the odd scattering matrix, such that:

  $$\mathbb{S}_o^{g'\to g}=diag\,\left(t_1\sigma^{g'\to g}_{s,1},t_3\sigma^{g'\to g}_{s,3},\dots\right),$$

[1] ie: for $1\leq g\leq G$, $m\in\mathcal{I}_e$, $(\phi^g_m)|_{\partial\mathcal{R}}=0$.

where $\sigma_{s,m}^{g'\to g}$ are the P_N moments of the macroscopic scattering cross sections from energy group g' to energy group g.

- $(\mathbb{M}_f)_{g,g'} := \chi^g \mathbb{M}_f^{g'} \in \mathbb{R}^{\hat{N}\times\hat{N}}$ is such that $\mathbb{M}_f^{g'}\phi^{g'} = (\underline{\nu\sigma}_f^{g'}\,\phi_0^{g'}, 0, \ldots)^T$ where the coefficient $\underline{\nu\sigma}_f^{g'}$ is the product of the number of neutrons emitted per fission times the macroscopic fission cross section; and the coefficient χ_g is the fission spectrum of energy group g.

Above, $t_l = \dfrac{\alpha_l^2}{(2l+1)}$ where α_l represents some normalisation factor such that: $\alpha_0 = 1$, $\alpha_l = \dfrac{4l^2-1}{l\alpha_l - 1}$ [29]. The coefficients of the matrices $\mathbb{T}_{e,o}$, $\mathbb{M}_f$ are supposed to be such that:

$$\left\{\begin{array}{ll}
\forall\, g,\, g' \in \mathcal{I}_G, \forall\, m \in \mathcal{I}_{e,o}: & \\
\quad (\sigma_{r,m}^g, \sigma_{s,m}^{g'\to g}, \underline{\nu\sigma}_f^g) \in \mathcal{P}W^{1,\infty}(\mathcal{R}) \times L^\infty(\mathcal{R}) \times L^\infty(\mathcal{R}). & (0)\\
\exists\, (\sigma_{r,(e,o)})_*,\, (\sigma^{r,(e,o)})^* > 0 \,|\, \forall\, g \in \mathcal{I}_G,\, \forall\, m \in \mathcal{I}_{e,o}: & \\
\quad (\sigma_{r,(e,o)})_* \le t_m \sigma_{r,m}^g \le (\sigma^{r,(e,o)})^* \text{ a.e. in } \mathcal{R}. & (i)\\
\exists\, (\underline{\nu\sigma}_f)^* > 0 \text{ such that } \forall\, g \in \mathcal{I}_G,\, 0 \le & \\
\quad \underline{\nu\sigma}_f^g \le (\underline{\nu\sigma}_f)^* \text{ a.e. in } \mathcal{R} \text{ and } \exists\, g' \,|\, \underline{\nu\sigma}_f^{g'} \ne 0. & (ii)\\
\exists\, 0 < \varepsilon < \dfrac{1}{G-1} \,|\, \forall\, m \in \mathcal{I}_{e,o},\, \forall\, g, g' \in \mathcal{I}_G, g' \ne g\,, |\sigma_{s,m}^{g\to g'}| \le \varepsilon \sigma_{r,m}^g \text{ a.e. in } \mathcal{R}. & (iii)
\end{array}\right. \tag{6.1.7}$$

Different homogenization steps allow to obtain these coefficients on the physical partition $\cup_{i=1}^{N_\mathcal{R}} \overline{\mathcal{R}_i}$. It happens that the coefficient $\underline{\nu\sigma}_f^g$ vanishes in some regions. Note that assumptions given in (6.1.7) help to validate the model. In particular, assumption (6.1.7)-(iii) is valid while modelling the core of a pressurized water reactor: the scattering cross-sections are weaker than the removal cross-sections of an order $0 < \varepsilon << 1$. Thus, the matrices $(\mathbb{T}_{e,o})^T$ are strictly diagonally dominant matrices: they are invertible.

Let us set $\mathbb{D} = \mathbb{H}^T\,\mathbb{T}_o^{-1}\,\mathbb{H}$. Problem (6.1.5) can be written as a set of coupled primal diffusion-like equations with single unknown $\phi \in \underline{V}$:

$$\text{Find } \phi \text{ such that } \; -\operatorname{div}(\mathbb{D}\,\nabla\,\phi) + \mathbb{T}_e\,\phi = S_f. \tag{6.1.8}$$

In the same way, Problem 6.1.6 can be written as:

$$\text{Find } (\lambda, \phi) \in \text{ such that } \; -\operatorname{div}(\mathbb{D}\,\nabla\,\phi) + \mathbb{T}_e\,\phi = \lambda^{-1}\mathbb{M}_f\phi. \tag{6.1.9}$$

Let us set $\alpha_{r,(e,o)} = \dfrac{(\sigma_{r,(e,o)})^*}{(\sigma_{r,(e,o)})_*} > 1$ and $\alpha_{s,(e,o)} := (G-1)\max\limits_{m\in\mathcal{I}_{e,o}}\max\limits_{g\ne g'\in\mathcal{I}_G}\sup\limits_{\vec{x}\in\mathcal{R}}\dfrac{|\sigma_{s,m}^{g'\to g}(\vec{x})|}{\sigma_{r,m}^g(\vec{x})}$. We have the following properties [A12, §3]:

Proposition 6.1.1 (`[A12, Proposition 3 and 4]`). *Suppose that $\alpha_{s,e} < (\alpha_{r,e})^{-1}$. The matrix $\mathbb{T}_e$ is positive definite:*

$$\forall X \in \mathbb{R}^{\hat{N}\times G} \quad (\mathbb{T}_e X|X) \ge \tau_e \|X\|^2 \text{ where } \tau_e = (\sigma_{r,e})_*\,(1 - \alpha_{r,e}\alpha_{s,e})\,. \tag{6.1.10}$$

Suppose that $\alpha_{s,o} < (\alpha_{r,o}+1)^{-1}$, the matrix $\mathbb{T}_o^{-1}$ is positive definite:

$$\forall X \in \mathbb{R}^{\hat{N}\times G} \quad (\mathbb{T}_o^{-1} X|X) \ge \tau_o \|X\|^2 \text{ where } \tau_o = \frac{1}{(\sigma_{r,o})^*}\left(1 - \frac{\alpha_{r,o}\alpha_{s,o}}{1-\alpha_{s,o}}\right). \tag{6.1.11}$$

We will study the discretization of the primal formulations (6.1.8) and (6.1.9) with discontinuous Galerkin finite elements in the last Section 6.7. Before in Sections 6.2-6.6, we focus on the discretization of the mixed formulations (6.1.5) and (6.1.6) in the case where $\hat{N} = 1$, and $G = 1$, i.e. for the neutron diffusion model, that we describe in the next paragraph.

6.1.4 Neutron diffusion model

Consider the SP_1 case, i.e. $\hat{N} = 1$.The pairs $(\mathbf{p}_g(\mathbf{x}), \phi_g(\mathbf{x}))_{g=1}^G$ satisfy:

$$\left\{ \begin{array}{rcl} \sigma_{r,1}^g \, \mathbf{p}_g + \dfrac{1}{3} \nabla \, \phi_g & = & \displaystyle\sum_{g' \neq g} \sigma_{s,1}^{g' \to g} \, \mathbf{p}_{g'}; \\ \\ \operatorname{div} \mathbf{p}_g + \sigma_{r,0}^g \, \phi_g & = & \lambda \dfrac{\chi^g}{4\pi} \displaystyle\sum_{g' \in \mathcal{I}_G} \nu^{g'} \sigma_f^{g'} \phi_{g'} + \sum_{g' \neq g} \sigma_{s,0}^{g' \to g} \, \phi_{g'}, \end{array} \right. \tag{6.1.12}$$

Furthermore, if we neglect the terms containing effective cross sections of order 1 in (6.1.12), we obtain the multi-group diffusion equations:

$$\left\{ \begin{array}{rcl} (D^g)^{-1} \, \mathbf{p}_g + \nabla \, \phi_g & = & 0 \text{ where } (D^g)^{-1} = 3 \, \sigma_t^g. \\ \operatorname{div} \mathbf{p}_g + \sigma_{r,0}^g \, \phi_g & = & \lambda \dfrac{\chi^g}{4\pi} \displaystyle\sum_{g'} \nu^{g'} \sigma_f^{g'} \phi_{g'} + \sum_{g' \neq g} \sigma_{s,0}^{g' \to g} \, \phi_{g'}, \end{array} \right. \tag{6.1.13}$$

In the single-group case ($G = 1$), the sum on g' is reduced to a single term and we have: $\chi^0 = 1$. Remove the superscript g from the notation. The pair $(\mathbf{p}, \phi)$ then satisfies the one-group diffusion equation:

$$\left\{ \begin{array}{rcl} D^{-1} \mathbf{p} + \nabla \, \phi & = & 0 \\ \operatorname{div} \, \mathbf{p} + \Sigma_a \, \phi & = & \lambda \underline{\nu \Sigma}_f \, \phi \end{array} \right. , \tag{6.1.14}$$

where $\Sigma_a = \sigma_{r,0}$ is the macroscopic absorption cross-section, and $\underline{\nu \Sigma}_f = \frac{1}{4\pi} \nu \sigma_f$ is the fission yield times the macroscopic fission cross section. Notice that $\operatorname{div} \mathbf{p} = - \operatorname{div} D \nabla \, \phi$, so that problem (6.1.14) is equivalent to the following primal problem:

$$- \operatorname{div} D \nabla \, \phi + \Sigma_a \, \phi = \lambda \underline{\nu \Sigma}_f \, \phi. \tag{6.1.15}$$

The diffusion coefficient, denoted by D, can be evaluated in two ways:

- Using $P_{N=1}$ approximation: $D^{-1} = 3(\sigma_t - \sigma_{s,1})$.
- Under some additional physical assumptions: i.e. the medium is homogeneous, the flux variations are small, and the flux spatial dependence is linear, computing the neutrons flux across a small surface yields Fick's law [80] with another coefficient: $D^{-1} = 3 \, \sigma_t$.

The $P_{N=1}$ approximation is more accurate. In that case, D is usually called the diffusion coefficient with transport correction. On the other hand, it is more costly, since one must derive $\sigma_{s,1}(\mathbf{x})$.

Equations (6.1.14) and (6.1.15) correspond to eigenproblems. Letting $\frac{1}{\lambda} = k_{eff}$, we look for $(k_{eff})^{-1}$, respectively ϕ, the *smallest* eigenvalue, resp. a related eigenflux.

Consider problem (6.1.14) in a bounded domain $\mathcal{R} \subset \mathbb{R}^d$, $d = 2$ or 3. Different boundary conditions can be taken into account, such as: zero flux: $\phi = 0$; reflection: $\mathbf{p} \cdot \mathbf{n} = 0$; albedo: $\mathbf{p} \cdot \mathbf{n} = \alpha \, \phi$, with $\alpha > 0$ (called vacuum when $\alpha = \frac{1}{2}$): $\mathbf{p} \cdot \mathbf{n} = \frac{1}{2} \, \phi$. Concerning problem (6.1.15), the first three conditions correspond respectively to:

- a *Dirichlet* boundary condition: $\phi = g_D$, with $g_D = 0$;
- a *Neumann* boundary condition: $\mathbf{p} \cdot \mathbf{n} = D \nabla \, \phi \cdot \mathbf{n} = g_N$, with $g_N = 0$;
- a *Robin* boundary condition: $\mathbf{p} \cdot \mathbf{n} + \alpha \, \phi = D \nabla \, \phi \cdot \mathbf{n} + \alpha \, \phi = g_R$, with $g_R = 0$.

The data $(D, \Sigma_a, \underline{\nu \Sigma}_f)$ are numerical data that are obtained from previous computations. They are piecewise polynomial on the partition $\{\mathcal{R}_i\}_{i \in \mathcal{I}_\mathcal{R}}$. To address the numerical analysis of problem (6.1.14),

we consider assumption given (6.1.7), which now read:

$$\left\{\begin{array}{l} (D,\Sigma_a,\underline{\nu\Sigma}_f)\in\mathbb{L}^\infty_{sym}(\mathcal{R})\times L^\infty(\mathcal{R})\times L^\infty(\mathcal{R})\,,\\ \exists D_*,D^*>0,\ \forall\mathbf{z}\in\mathbb{R}^d,\ D_*\|\mathbf{z}\|^2\le(D\mathbf{z},\mathbf{z})\le D^*\|\mathbf{z}\|^2 \text{ a.e. in } \mathcal{R}\,,\\ \exists(\Sigma_a)_*,(\Sigma_a)^*>0,\ 0<(\Sigma_a)_*\le\Sigma_a\le(\Sigma_a)^* \text{ a.e. in } \mathcal{R}\,,\\ 0\le\underline{\nu\Sigma}_f \text{ a.e. in } \mathcal{R}\,,\ \underline{\nu\Sigma}_f\neq 0. \end{array}\right. \tag{6.1.16}$$

In particular, it can happen that $\underline{\nu\Sigma}_f$ vanishes on some regions.

We will use two domain decomposition methods. The partition of the domain decomposition, that we will denote by $\{\widetilde{\mathcal{R}}_i\}_{i\in\tilde{\mathcal{I}}_\mathcal{R}}$, $\tilde{\mathcal{I}}_\mathcal{R}=\{1,\dots,\tilde{N}_\mathcal{R}\}$, may be independent from the physical partition of the materials $\{\widetilde{\mathcal{R}}_i\}_{i\in\tilde{\mathcal{I}}_\mathcal{R}}$. In other word, $\{\widetilde{\mathcal{R}}_i\}_{i\in\tilde{\mathcal{I}}_\mathcal{R}}\neq\{\mathcal{R}_i\}_{i\in\mathcal{I}_\mathcal{R}}$. We will then use the notation $\mathcal{P}\cdot$ (resp. $\tilde{\mathcal{P}}\cdot$) for functional spaces that are piecewise regular on $\{\mathcal{R}_i\}_{i\in\mathcal{I}_\mathcal{R}}$ (resp. $\{\widetilde{\mathcal{R}}_i\}_{i\in\tilde{\mathcal{I}}_\mathcal{R}}$). Note that if here we study spatial domain decomposition, one can also perform parallel in time simulations of the neutron transport equation, cf. [30, 31, 200]. Another approach to improve computation time inline is to use reduced basis method [14, 238].

As we look for the smallest eigenvalue k_{eff} of Equation (6.1.14), the solution $(\mathbf{p},\phi,k_{eff})$ can be computed by the inverse power iteration algorithm, which is detailed in Appendix A.1.

We refer to [A10, P6, A11] for more details on the multigroup SP_N and diffusion neutron equations and the general algorithm to solve them. From now on, we focus on the study of the neutron diffusion equation.

6.2 Neutron diffusion, source problem [A11]

First, we consider equations (6.1.14), with a source term S_f, namely we focus on the linear system solver step of the above algorithm. We split the boundary of the domain $\mathcal{R}$ into three disjoint, open parts such that: $\partial\mathcal{R}=\overline{\Gamma}_D\cup\overline{\Gamma}_N\cup\overline{\Gamma}_R$. The neutron mixed diffusion equation reads:

$$\left\{\begin{array}{rcll} \text{Find } (\mathbf{p},\phi) \text{ such that} & & & \\ D^{-1}\,\mathbf{p}+\nabla\,\phi &=& 0 & \text{in } \mathcal{R}\\ \operatorname{div}\mathbf{p}+\Sigma_a\phi &=& S_f & \text{in } \mathcal{R}\\ \phi &=& 0 & \text{on } \Gamma_D\\ \mathbf{p}\cdot\mathbf{n} &=& 0 & \text{on } \Gamma_N\\ -\mathbf{p}\cdot\mathbf{n}+\alpha\phi &=& 0 & \text{on } \Gamma_R \end{array}\right. . \tag{6.2.1}$$

Note that the current $\mathbf{p}$ is such that $\mathbf{p}\in\mathbf{L}^2(\mathcal{R})$. Define $H^1_{0,\Gamma_D}(\mathcal{R}):=\{\psi\in H^1(\mathcal{R})\,|\,\psi_{|\Gamma_D}=0\}$. Recall that $(\mathbf{p},\phi)$ is governed by (6.2.1): as $\mathbf{p}\in\mathbf{L}^2(\mathcal{R})$, we deduce that $\phi\in H^1_{0,\Gamma_D}(\mathcal{R})$. Furthermore, we can rewrite equivalently equations (6.2.1) with the flux $\phi\in H^1_{0,\Gamma_D}(\mathcal{R})$ as the only variable, governed by:

$$\left\{\begin{array}{rcll} \text{Find } \phi \text{ such that} & & & \\ -\operatorname{div}(\,D\,\nabla\,\phi\,)+\Sigma_a\phi &=& S_f & \text{in } \mathcal{R},\\ D\partial_n\phi &=& 0 & \text{on } \Gamma_N,\\ D\partial_n\phi+\alpha\phi &=& 0 & \text{on } \Gamma_R. \end{array}\right. \tag{6.2.2}$$

Then, we recover $\mathbf{p}$ using the relation $\mathbf{p}=-D\,\nabla\phi$. About Eqs. (6.2.1) and (6.2.2), one has the elementary results below.

Theorem 6.2.1. *There exists a unique solution $\phi\in H^1_{0,\Gamma_D}(\mathcal{R})$ to Problem* (6.2.2).

Define $\mathbf{H}_{0,\Gamma_N}(\operatorname{div};\mathcal{R}):=\{\mathbf{q}\in\mathbf{H}(\operatorname{div};\mathcal{R})\,|\,\mathbf{q}\cdot\mathbf{n}_{|\Gamma_N}=0\}$. By construction, we have that $\mathbf{p}\in\mathbf{H}_{0,\Gamma_N}(\operatorname{div};\mathcal{R})$ and also $\mathbf{p}\cdot\mathbf{n}_{|\Gamma_R}\in L^2(\Gamma_R)$. So, we introduce the natural functional space and related

norm for the currents

$$\mathbf{Q} := \{\mathbf{q} \in \mathbf{H}_{0,\Gamma_N}(\mathrm{div};\mathcal{R}) \,|\, \mathbf{q}\cdot\mathbf{n}_{|\Gamma_R} \in L^2(\Gamma_R)\},\ \|\mathbf{q}\|_{\mathbf{Q}} := \left(\|\mathbf{q}\|^2_{\mathbf{H}(\mathrm{div};\mathcal{R})} + \int_{\Gamma_R}(\mathbf{q}\cdot\mathbf{n})^2\right)^{1/2}.$$

Since (6.2.1) is equivalent to (6.2.2), we infer from Theorem 6.2.1 the second result below.

Theorem 6.2.2. *There exists a unique* $(\mathbf{p},\phi) \in \mathbf{Q}\times H^1_{0,\Gamma_D}(\mathcal{R})$ *to Problem* (6.2.1)*, such that* ϕ *satisfies Problem* (6.2.2)*.*

6.2.1 Variational formulation

Let us set $\mathbf{X} = \mathbf{Q}\times L^2(\mathcal{R})$, endowed with the norm: $\|(\mathbf{q},\psi)\|_{\mathbf{X}} = \left(\|\mathbf{q}\|^2_{\mathbf{Q}} + \|\psi\|^2_{L^2(\mathcal{R})}\right)^{\frac{1}{2}}$. Consider the bilinear form $c : \mathbf{X}\times\mathbf{X}\to\mathbb{R}$ such that for all $(\,(\mathbf{q},\psi),(\mathbf{r},\varphi)\,) \in \mathbf{X}\times\mathbf{X}$

$$c((\mathbf{q},\psi),(\mathbf{r},\varphi)) = \int_{\mathcal{R}}\left(-D^{-1}\mathbf{q}\cdot\mathbf{r} + \varphi\ \mathrm{div}\,\mathbf{q} + \psi\ \mathrm{div}\,\mathbf{r} + \Sigma_a\psi\,\varphi\right) - \int_{\Gamma_R}\frac{1}{\alpha}(\mathbf{q}\cdot\mathbf{n})(\mathbf{r}\cdot\mathbf{n}). \tag{6.2.3}$$

The bilinear form $c(\cdot,\cdot)$ is T-coercive.

Theorem 6.2.3 (`[A11, Theorem 3]`)**.** *Assume that the data* D *and* Σ_a *satisfy assumptions given in* (6.1.7)*. Then it exists* $T\in\mathcal{L}(\mathbf{X})$*, bijective, it exists* $\alpha_T > 0$ *such that for all* $(\mathbf{q},\psi)\in\mathbf{X}$*,*

$$|c((\mathbf{q},\psi),T(\mathbf{q},\psi))| \geq \alpha_T\|(\mathbf{q},\psi)\|^2_{\mathbf{X}}.$$

Proof. Consider $T\in\mathcal{L}(\mathbf{X})$ s.t. $T((\mathbf{q},\psi)) = (\mathbf{q}^\star,\psi^\star) = (-\mathbf{q},\frac{1}{2}\psi + \frac{1}{2}(\Sigma_a)^{-1}\,\mathrm{div}\,\mathbf{q})$. We obtain:

$$\begin{array}{rcl} c((\mathbf{q}^\star,\psi^\star),(\mathbf{q},\psi)) & \geq & \min\left((D^*)^{-1},(2(\Sigma_a)^*)^{-1},\alpha^{-1},\dfrac{(\Sigma_a)_*}{2}\right)\|(\mathbf{q},\psi)\|^2_{\mathbf{X}},\\ \|T((\mathbf{q},\psi))\|^2_{\mathbf{X}} = \|(\mathbf{q}^\star,\psi^\star)\|^2_{\mathbf{X}} & \leq & \max\left(1+\frac{1}{2}((\Sigma_a)_*)^{-2},\frac{1}{2}\right)\|(\mathbf{q},\psi)\|^2_{\mathbf{X}}. \end{array}$$

The injectivity of operator T is clear: given $(\mathbf{q}^\star,\psi^\star)\in\mathbf{X}$, choosing $(\mathbf{q},\psi) = (-\mathbf{q}^\star, 2\psi^\star + (\Sigma_a)^{-1}\,\mathrm{div}\,\mathbf{q}^\star$ yields $T((\mathbf{q},\psi)) = (\mathbf{q}^\star,\psi^\star)$. Hence, the operator $T\in\mathcal{L}(\mathbf{X})$ is bijective. ◇

Let consider the linear and continuous form $\ell_{S_f}\in\mathcal{L}(\mathbf{X})$ such that for all $(\mathbf{q},\psi)\in\mathbf{X}$, $\ell_{S_f}(\,(\mathbf{q},\psi)\,) = \displaystyle\int_{\mathcal{R}} S_f\,\psi$. The mixed variational formulation governing $(\mathbf{p},\phi)$ reads:

$$\text{Find } (\mathbf{p},\phi)\in\mathbf{X} \text{ such that for all } (\mathbf{q},\psi)\in\mathbf{X}\quad c(\,(\mathbf{p},\phi),(\mathbf{q},\psi)\,) = \ell_{S_f}(\,(\mathbf{q},\psi)\,). \tag{6.2.4}$$

According to the abstract T-coercivity theory, one has the

Theorem 6.2.4. *Assume that the data* D *and* Σ_a *satisfy assumptions given in* (6.1.7)*. Then, there exists a unique solution* $(\mathbf{p},\phi)\in\mathbf{X}$ *of the mixed variational formulation* (6.2.4)*, and moreover this solution satisfies problem (6.2.1).*

6.2.2 Discretization

We recall below the definition of the Raviart-Thomas-Nédélec (or RTN) finite element [213, 204]. Let $(K_\ell)_{1\leq\ell\leq L}$ be a conforming mesh, or triangulation, of $\overline{\mathcal{R}}$ made of parallelepipeds (a mesh, or triangulation, is said to be conforming if in every K_ℓ, D and Σ_a are smooth). Let $P(K_\ell)$ be the set of polynomials defined over K_ℓ. For integer values $l,m,p\geq 0$, we consider the following subspace of $P(K_\ell)$:

$$Q_{l,m,p}(K_\ell) = \left\{ q(x,y,z)\in P(K_\ell)\,|\, q(x,y,z) = \sum_{e,j,k=0}^{l,m,p} a_{e,j,k}\,x^e\,y^j\,z^k,\ a_{e,j,k}\in\mathbb{R}\right\}.$$

For integer $k \geq 0$, let us set $k' = k+1$ and introduce the vector polynomial space:

$$\mathbf{D}_k(K_\ell) = [Q_{k',k,k}(K_\ell) \times \mathbf{0} \times \mathbf{0}] \oplus [\mathbf{0} \times Q_{k,k',k}(K_\ell) \times \mathbf{0}] \oplus [\mathbf{0} \times \mathbf{0} \times Q_{k,k,k'}(K_\ell)].$$

We can now define the $\mathrm{RTN}_{[k]}$ finite element subspace of $\mathbf{H}(\mathrm{div}, \mathcal{R}) \times L^2(\mathcal{R})$:

$$\begin{aligned} \mathbf{Q}_h^k &= \left\{ \mathbf{q} \in \mathbf{H}(\mathrm{div}, \mathcal{R}) \,|\, \forall \ell \in \{1, \ldots, L\},\, \mathbf{q}_{|K_\ell} \in \mathbf{D}_k(K_\ell) \right\}, \\ L_h^k &= \left\{ \psi \in L^2(\mathcal{R}) \,|\, \forall \ell \in \{1, \ldots, L\},\, \psi_{|K_\ell} \in Q_{k,k,k}(K_\ell) \right\}. \end{aligned} \tag{6.2.5}$$

For later use, we denote π^0 the $L^2(\mathcal{R})$ orthogonal projector on its subspace L_h^0. By construction, it holds $\mathrm{range}(\pi^0) = L_h^0$ where π^0 is defined by:

$$\forall \psi \in L^2(\mathcal{R}),\, \forall \psi_h \in L_h^0, \quad (\pi^0 \psi - \psi, \psi_h)_{L^2(\mathcal{R})} = 0. \tag{6.2.6}$$

According to [146, Proposition 1.135]:

$$\begin{array}{lll} \forall z \in L^2(\mathcal{R}), & \|z - \pi^0 z\|_{L^2(\mathcal{R})} & \lesssim \|z\|_{L^2(\mathcal{R})}, \\ \forall z \in \mathcal{P}H^1(\mathcal{R}), & \|z - \pi^0 z\|_{L^2(\mathcal{R})} & \lesssim h\, \|z\|_{\mathcal{P}H^1(\mathcal{R})}, \\ \forall z \in \mathcal{P}W^{1,\infty}(\mathcal{R}), & \|z - \pi^0 z\|_{L^\infty(\mathcal{R})} & \lesssim h\, \|z\|_{\mathcal{P}W^{1,\infty}(\mathcal{R})}. \end{array} \tag{6.2.7}$$

For the last two inequalities, the result holds provided that the triangulations are conforming with respect to the partition, namely for all triangulations, for all elements K of a triangulation, it holds that there exists $i \in \mathcal{I}_\mathcal{R}$ such that $K \subset \overline{\mathcal{R}_i}$. Similar results hold on subsets of $\mathcal{R}$.

As required, it holds $\mathrm{div}\, \mathbf{Q}_h^k \subset L_h^k$ and $L_h^0 \subset L_h^k$. We recall that for any $\mathbf{q}$ in $\mathbf{H}(\mathrm{div}, \mathcal{R})$, its $\mathrm{RTN}_{[k]}$-interpolant $\mathbf{q}_R^k \in \mathbf{Q}_h^k$ satisfies:

$$\forall \psi_h \in L_h^k,\ b(\mathbf{q} - \mathbf{q}_R^k, \psi_h) = 0. \tag{6.2.8}$$

In addition thanks to the commuting diagram property, cf. [50, §2.5.2], it holds

$$\forall \mathbf{q} \in \mathbf{H}(\mathrm{div}, \mathcal{R}),\ \mathrm{div}\, \mathbf{q}_R^0 = \pi^0(\mathrm{div}\, \mathbf{q}). \tag{6.2.9}$$

Let $\mathbf{q} \in \mathbf{H}^r(\mathcal{R})$, such that $\mathrm{div}\, \mathbf{q} \in H^s(\mathcal{R})$, $0 < r, s < r_{\max}$. According to [34, Lemma 3.3]:

$$\begin{aligned} \|\mathbf{q} - \mathbf{q}_R^0\|_{\mathbf{L}^2(\mathcal{R})} &\lesssim \left(h^r |\mathbf{q}|_{\mathbf{H}^r(\mathcal{R})} + h\, \|\mathrm{div}\, \mathbf{q}\|_{L^2(\mathcal{R})} \right), \\ \|\mathrm{div}(\mathbf{q} - \mathbf{q}_R^0)\|_{L^2(\mathcal{R})} &\lesssim h^s |\mathrm{div}\, \mathbf{q}|_{H^s(\mathcal{R})}. \end{aligned} \tag{6.2.10}$$

Similar results hold on subsets of $\mathcal{R}$, provided the discretizations are conforming.

Remark 6.2.1.

- *Note that the Neumann boundary condition corresponds to essential boundary conditions for the discrete vector fields, so that we will have to eliminate the degrees of freedom on Γ_N if it exists. By construction, for all $\mathbf{q} \in \mathbf{Q}_h^k$, one has automatically that $\mathbf{q} \cdot \mathbf{n}_{|\partial\mathcal{R}} \in L^2(\partial\mathcal{R})$, and it follows that $\mathbf{Q}_h^k \subset \mathbf{Q}$. In other words, the discretization $\mathbf{X}_h^k$ is conforming in $\mathbf{X}$. Moreover, we have $\mathrm{div}\, \mathbf{Q}_h^k \subset L_h^k$: consistency is ensured.*

- *If one chooses another discretization, all results presented hereafter hold provided the estimates (6.2.10) remain true. For instance, for the RTN$_{[k]}$ finite element defined on tetrahedral triangulations of $\overline{\mathcal{R}}$, cf. [50, §2.3.1]. To prove (6.2.10) in this case, one has simply to apply the results of [34, §3.2]. On the other hand, provided that the field $\mathbf{q}$ and its divergence are "smooth" in the sense that they belong to $\mathcal{P}H^{m+1}(\mathcal{R})$ for some integer $m \geq 0$, using the RTN$_{[m]}$ finite element one can recover interpolation estimates in $O(h^{m+1})$, cf. [50, §2.5.5]. For meshes made of affine elements such as tetrahedra or parallelepipeds, the approximation estimate (6.2.10-top) does not require the term with the divergence (see, e.g. [50], §2.5.1).*

The discrete variational formulation reads:

$$\text{Find } (\mathbf{p}_h, \phi_h) \in \mathbf{X}_h^k \text{ s.t. } \forall (\mathbf{q}_h, \psi_h) \in \mathbf{X}_h^k: \quad c((\mathbf{p}_h, \phi_h), (\mathbf{q}_h, \psi_h)) = \int_{\mathcal{R}} S_f \, \psi_h. \tag{6.2.11}$$

Problem (6.2.11) is well-posed, cf. [217, 146]. We prove it using the discrete T-coercity.

Theorem 6.2.5 (`[A2, Theorem 4.5]`). *It exists $T_h \in \mathcal{L}(\mathbf{X}_h^k)$, bijective, it exists α_{T_h} such that for all for all $(\mathbf{q}_h, \psi_h) \in \mathbf{X}_h^k$,*

$$|c((\mathbf{q}_h, \psi_h), T_h(\mathbf{q}_h, \psi_h))| \geq \alpha_{T_h} \, \|(\mathbf{q}_h, \psi_h)\|_{\mathbf{X}}^2. \tag{6.2.12}$$

Proof. We follow the proof of Theorem 6.2.3, using consistency property of the finite element spaces $\mathbf{Q}_h^k$ and L_h^k. One can remark that if Σ_a is piecewise-constant, $\frac{1}{2}(\Sigma_a)^{-1} \operatorname{div} \mathbf{p}_h$ is automatically in L_h. Otherwise, we project $(\Sigma_a)^{-1}$ on the piecewise-constant functions. One chooses $T_h(\,(\mathbf{q}_h, \psi_h)\,) := (-\mathbf{p}_h, \frac{1}{2}\phi_h + \frac{1}{2}\pi^0((\Sigma_a)^{-1}) \operatorname{div} \mathbf{p}_h)$. According to Equation (6.2.7), $\|(\Sigma_a)^{-1} - \pi^0((\Sigma_a)^{-1})\|_{L^\infty(\mathcal{R})} \lesssim h$, which allows us a uniform discrete inf-sup condition. ⋄

6.2.3 A priori estimates

Convergence result can be found in [217] (see Theorem 13.2, p. 582, and (6.19-6.20) pp. 553-554). It makes use of local estimates.

Theorem 6.2.6 (`[A11, Theorem 4]`). *The discrete solution converges to the exact solution as h goes to zero. Moreover, if $\phi \in H^\mu(\mathcal{R})$, $\mathbf{p} \in \mathbf{H}^p(\mathcal{R})$ and $\operatorname{div} \mathbf{p} \in H^{p'}(\mathcal{R})$ with $f, p, p' \geq 1$, it exists a constant $C > 0$ independent of h such that*

$$\|\mathbf{p} - \mathbf{p}_h\|_{\mathbf{H}(\operatorname{div},\mathcal{R})} + \|\phi - \phi_h\|_{L^2(\mathcal{R})} \leq C \, h^\alpha \left(|\phi|_{H^\mu(\mathcal{R})} + |\mathbf{p}|_{H^p(\mathcal{R})} + |\operatorname{div} \mathbf{p}|_{H^{p'}(\mathcal{R})} \right), \tag{6.2.13}$$

where $\alpha = \min(k, \mu, p, p')$.

Several illuminating comments can be made about these convergence results:

- One has always $\phi \in H^1(\mathcal{R})$.
- On the other hand, if S_f belongs to $L^2(\mathcal{R})$ (minimal regularity), then $\operatorname{div} \mathbf{p} = S_f - \Sigma_a \phi \in L^2(\mathcal{R})$ according to (6.2.2).
- Suppose that $D = 1$ over $\mathcal{R}$: as a consequence, the relation between ϕ and $\mathbf{p}$ reads $\mathbf{p} = -\nabla \phi$. It follows that the regularity of the vector field $\mathbf{p}$ is always equal to that of the scalar field ϕ *minus one*: $p = \mu - 1$.
- Suppose still that $D = 1$ over $\mathcal{R}$: then $\mathbf{p} \in \mathbf{L}^2(\mathcal{R})$, $\mathbf{curl}\, \mathbf{p} = 0$ and $\operatorname{div} \mathbf{p} \in L^2(\mathcal{R})$. In addition, the boundary conditions yield $\mathbf{p} \times \mathbf{n}_{|\Gamma_D} = 0$ (because $\phi_{|\Gamma_D} = 0$), $\mathbf{p} \cdot \mathbf{n}_{|\Gamma_N} = 0$, and $\mathbf{p} \cdot \mathbf{n}_{|\Gamma_R} \in H^{1/2}(\Gamma_R)$. According to [166, 167], one has $\mathbf{p} \in \mathbf{H}^p(\mathcal{R})$, with $p > 1/4$ in all configurations, and $p > 1/2$ provided that Γ_D is perpendicular to Γ_N and Γ_R.
- The estimate (6.2.13) can be generalized in the following way. Following the proofs given in [217], one finds that the regularity assumptions on ϕ, $\mathbf{p}$ and $\operatorname{div} \mathbf{p}$ need only to be local. In other words, only *piecewise regularity* of the solution is required for (6.2.13) to hold, which can be of great importance if the coefficients D and Σ_a are themselves piecewise constant coefficients.

Consider the low regular case, i.e. $S_f \in H^\mu(\mathcal{R})$ with $\mu \in]0, r_{\max}[$, $r_{\max} < 1/2$:

Theorem 6.2.7 (`[A2, Theorem 4.7, Theorem 4.10]`). *Under the assumptions of Proposition 4.4.1, it holds, with $r_{\max} < 1/2$: $\forall \mu \in]0, r_{\max}[$, $\forall S_f \in H^\mu(\mathcal{R})$,*

$$\|\mathbf{p} - \mathbf{p}_h\|_{\mathbf{H}(\operatorname{div},\mathcal{R})} + \|\phi - \phi_h\|_{L^2(\mathcal{R})} \lesssim h^\mu \, \|S_f\|_{H^\mu(\mathcal{R})}, \tag{6.2.14}$$

$$\|\phi - \phi_h\|_{L^2(\mathcal{R})} \lesssim h^{2\mu} \, \|S_f\|_{H^\mu(\mathcal{R})}. \tag{6.2.15}$$

Remark 6.2.2. *In particular, for "smooth data" S_f, i.e. $S_f \in H^{r_{\max}}(\mathcal{R})$, one expects a convergence rate at least in $h^{r_{\max}-\eta}$ for $\eta > 0$ arbitrary small: by a slight* abuse of notation *there and in the sequel, we shall write $h^{r_{\max}}$. Also, the previous analysis can be extended to the case where $r_{\max}$ is in $[1/2, 1]$ and $\mu < r_{\max}$ (or $\mu \leq 1$ if $r_{\max} = 1$). Furthermore, for a "smooth" solution, one may recover a convergence rate like $O(h^{m+1})$ for an $RTN_{[m]}$ discretization of order $m \geq 0$.*

6.2.4 Explicit T-coercivity

Let T be defined as in the proof of Theorem 6.2.3: $T((\mathbf{q},\psi)) = (-\mathbf{q}, \frac{1}{2}\psi + \frac{1}{2}(\Sigma_a)^{-1}\operatorname{div}\mathbf{q})$. Since T is bijective, the variational formulation (6.2.4) is equivalent to:

$$\text{Find } (\mathbf{p},\phi) \in \mathbf{X} \text{ such that for all } (\mathbf{q},\psi) \in \mathbf{X} \quad c'((\mathbf{p},\phi), T((\mathbf{q},\psi))) = \ell_{S_f}(T(\mathbf{q},\psi)). \tag{6.2.16}$$

By design, the form $c' : ((\mathbf{p},\phi),(\mathbf{q},\psi)) \mapsto c((\mathbf{p},\phi), T((\mathbf{q},\psi)))$ is coercive. Using $T((\mathbf{q},\psi)) = (-\mathbf{q}, \frac{1}{2}\psi + \frac{1}{2}(\Sigma_a)^{-1}\operatorname{div}\mathbf{q})$, one can compute its expression:

$$\begin{aligned} c'((\mathbf{p},\phi),(\mathbf{q},\psi)) = &\int_{\mathcal{R}} D^{-1}\mathbf{p}\cdot\mathbf{q} + \frac{1}{2}\int_{\mathcal{R}} (\Sigma_a)^{-1}\operatorname{div}\mathbf{p}\,\operatorname{div}\mathbf{q} + \frac{1}{\alpha}\int_{\Gamma_R} (\mathbf{p}\cdot\mathbf{n})\,(\mathbf{q}\cdot\mathbf{n}) \\ &-\frac{1}{2}\int_{\mathcal{R}} \phi\,\operatorname{div}\mathbf{q} + \frac{1}{2}\int_{\mathcal{R}} \psi\,\operatorname{div}\mathbf{p} + \frac{1}{2}\int_{\mathcal{R}} \Sigma_a\phi\psi. \end{aligned} \tag{6.2.17}$$

Let us set $\ell'_{S_f}(\mathbf{q},\psi) = \ell_{S_f}(T(\mathbf{q},\psi))$, we have:

$$\ell'_{S_f}(\mathbf{q},\psi) = \frac{1}{2}\int_{\mathcal{R}} S_f\left(\psi + (\Sigma_a)^{-1}\operatorname{div}\mathbf{q}\right).$$

We can write problem (6.2.16) with two equations as follows (with $\gamma = \frac{1}{2}$):

$$\left\{\begin{array}{l} \text{Find } (\mathbf{p},\phi) \in \mathbf{Q}\times L^2(\mathcal{R}) \text{ such that} \\ \\ \displaystyle\int_{\mathcal{R}} D^{-1}\mathbf{p}\cdot\mathbf{q} + \gamma\int_{\mathcal{R}} (\Sigma_a)^{-1}\operatorname{div}\mathbf{p}\,\operatorname{div}\mathbf{q} + \frac{1}{\alpha}\int_{\Gamma_R} (\mathbf{p}\cdot\mathbf{n})\,(\mathbf{q}\cdot\mathbf{n}) \\ \qquad -(1-\gamma)\int_{\mathcal{R}} \phi\,\operatorname{div}\mathbf{q} \;=\; \gamma\displaystyle\int_{\mathcal{R}} S_f(\Sigma_a)^{-1}\operatorname{div}\mathbf{q} \quad \forall \mathbf{q}\in\mathbf{Q} \\ \\ \gamma\displaystyle\int_{\mathcal{R}} \psi\,\operatorname{div}\mathbf{p} + \gamma\int_{\mathcal{R}} \Sigma_a\phi\psi \;=\; \gamma\int_{\mathcal{R}} S_f\psi \qquad\qquad \forall \psi\in L^2(\mathcal{R}) \end{array}\right. . \tag{6.2.18}$$

Recall the continuous problem (6.2.1):

$$\left\{\begin{array}{l} \text{Find } (\mathbf{p},\phi) \in \mathbf{Q}\times H^1_{0,\Gamma_D}(\mathcal{R}) \text{ such that} \\ D^{-1}\,\mathbf{p} + \nabla\,\phi \;=\; 0 \quad \text{in } \mathcal{R} \quad \text{(i)} \\ \operatorname{div}\mathbf{p} + \Sigma_a\phi \;=\; S_f \quad \text{in } \mathcal{R} \quad \text{(ii)} \\ -\mathbf{p}\cdot\mathbf{n} + \alpha\phi \;=\; 0 \quad \text{on } \Gamma_R \quad \text{(iii)} \end{array}\right. . \tag{6.2.19}$$

We can obtain the first equation of problem (6.2.18) by a stabilization method, with $\gamma = \frac{1}{2}$:

1. Multiply (6.2.19)-(i) by $\mathbf{q} \in \mathbf{Q}$ and integrate over $\mathcal{R}$, use (6.2.19)-(iii) to integrate the gradient term by parts.
2. Multiply (6.2.19)-(i) by $\gamma(\Sigma_a)^{-1}\operatorname{div}\mathbf{q}$, where $\mathbf{q} \in \mathbf{Q}$ and $\gamma > 0$, and integrate over $\mathcal{R}$.
3. Add these two equations together.

Remarks 6.2.1. *We observe that*

- *In the case where $\gamma = 0$, the first equation of* (6.2.18) *is independent of ϕ. We can calculate the unknown ϕ once $\mathbf{p}$ is evaluated.*

- *We can solve problem* (6.2.16) *directly. When we apply the explicit T-coercivity to solve magnetostatic systems, cf. Section 5.2, we obtain a pertubed approach. In the case of the Stokes problem, cf. Section 7.7, we need an initial guess for the pressure.*

6.3 Neutron diffusion, generalized eigenproblem [A2, 163]

6.3.1 Mixed formulation

We now consider the neutron diffusion eigenproblem (6.1.14). We recall that the approximation of eigenvalue problems has been studied among others by Osborn et al in [206, 18], and in particular by Boffi et al [51, 52, 49] when the eigenproblem is in a mixed form. In our case however, their theory does not ensure the spectral correctness of the approximation so we design a new proof to obtain this result. On the other hand, we can adapt the work of Boffi et al [53] to exhibit a convergence rate for the eigenvalues.

Consider the generalized eigenvalue problem (6.1.14) with the assumptions given in 6.1.7, supplemented with $\underline{\nu\Sigma}_f \in \mathcal{P}W^{1,\infty}(\mathcal{R})$.

$$\left\{\begin{array}{rcl} \text{Find } (\mathbf{p},\phi) \in \mathbf{H}(\mathrm{div},\mathcal{R}) \times L^2(\mathcal{R}) \text{ and } k_{eff} \text{ such that} & & \\ D^{-1}\mathbf{p} + \nabla\,\phi &=& 0 \\ \mathrm{div}\ \mathbf{p} + \Sigma_a\,\phi &=& \lambda \underline{\nu\Sigma}_f\,\phi \\ \phi_{|\partial\mathcal{R}} &=& 0 \\ k_{eff} &=& 1/\min_\lambda \lambda \end{array}\right. . \tag{6.3.1}$$

Let $0 \le \mu < r_{\max}$ be given, we introduce an operator B_μ associated to the source problem (6.2.2): given $f \in H^\mu(\mathcal{R})$, we call $B_\mu f = \phi \in H^1(\mathcal{R})$ the second component of the couple $(\mathbf{p},\phi)$ that solves (6.2.1) with $\Gamma_D = \partial\mathcal{R}$ and a source $S_f = \underline{\nu\Sigma}_f f$. Since $\underline{\nu\Sigma}_f$ belongs to $\mathcal{P}W^{1,\infty}(\mathcal{R})$, it holds $\|S_f\|_{H^\mu(\mathcal{R})} \lesssim \|f\|_{H^\mu(\mathcal{R})}$ because $\mu < \frac{1}{2}$. Hence, B_μ is a bounded operator from $H^\mu(\mathcal{R})$ to itself:

$$\|B_\mu f\|_{H^\mu(\mathcal{R})} \lesssim \|B_\mu f\|_{H^1(\mathcal{R})} = \|\phi\|_{H^1(\mathcal{R})} \lesssim \|S_f\|_{L^2(\mathcal{R})} \lesssim \|S_f\|_{H^\mu(\mathcal{R})} \lesssim \|f\|_{H^\mu(\mathcal{R})}\,; \tag{6.3.2}$$

we write $B_\mu \in \mathcal{L}(H^\mu(\mathcal{R}))$ for short. In addition, since the second component of the solution actually belongs to $H^1(\mathcal{R})$ with continuous dependence ($\|\phi\|_{H^1(\mathcal{R})} \lesssim \|f\|_{H^\mu(\mathcal{R})}$), it follows that B_μ is a compact operator. Denote by $\sigma(B_\mu)$ its spectrum. By construction, $\lambda^{-1} \in \sigma(B_\mu)$ if, and only if, λ is an eigenvalue of problem (6.1.15).

6.3.2 Estimates for the generalized eigenproblem

Consider the discretization of problem (6.3.1) with $\mathrm{RTN}_{[k]}$ finite elements. We call the discrete operator B_μ^h associated to the discrete source problem (6.2.11): given $f \in H^\mu(\mathcal{R})$, we call $B_\mu^h f$ the second component of the couple $(\mathbf{p}_h,\phi_h)$ that solves (6.2.11) with source $S_f = \underline{\nu\Sigma}_f f$, so that it holds: $\lim_{h\to 0} \|B_0 f - B_0^h f\|_{L^2(\mathcal{R})} = 0$ for all $f \in L^2(\mathcal{R})$. This property is the so-called pointwise convergence.

However, for a mixed formulation, the fact that the family $(B_0^h)_h$ converges pointwise towards the compact operator B_0 is not *sufficient* to guarantee that the family $(B_0^h)_h$ converges in operator norm towards B_0.

Convergence in operator norm

According to [206], proving that $\lim_{h\to 0} \|B_\mu - B_\mu^h\|_{\mathcal{L}(H^\mu(\mathcal{R}))} = 0$ for discrete approximants $(B_\mu^h)_h$ is a sufficient condition to obtain convergence of the eigenvalues. In order to ensure the convergence in operator norm of the family $(B_\mu^h)_h$ towards the compact operator B_μ, we need a technical assumption on the triangulations.

Definition 6.3.1. *A family of triangulations $(\mathcal{T}_h)_h$ is* regular$^+$ *if it satisfies:*

$$\exists 0 < \theta \leq 1,\ \forall h,\ h^{2-\theta} \lesssim \min_{K\in\mathcal{T}_h} diam(K). \tag{6.3.3}$$

It corresponds to a triangulation such that $h/\min_{K\in\mathcal{T}_h} diam(K) \lesssim h^{\theta-1}$, which may limit adaptive h-refinement.In particular, a quasi-uniform family of triangulations is *regular*$^+$(take $\theta = 1$ in (6.3.3)). For a *regular*$^+$ family, one has the following inverse inequality, whose proof follows [33, Section 2].

Lemma 6.3.1 (`[A2, Lemma 4.13]`)**.** *Let $\mu \in [0, 1/2[$. For a* regular$^+$ *family of triangulations, it holds:*

$$\forall h,\ \forall \psi_h \in L_h^k,\ \|\psi_h\|_{H^\mu(\mathcal{R})} \lesssim h^{-2\mu+\theta\mu}\|\psi_h\|_{L^2(\mathcal{R})}. \tag{6.3.4}$$

Theorem 6.3.1 (`[A2, Theorem 4.14]`)**.** *Under the assumptions of Proposition 4.4.1 with $r_{\max} < 1/2$ plus $\underline{\nu\Sigma}_f \in \mathcal{P}W^{1,\infty}(\mathcal{R})$, let $\mu \in [0, r_{\max}[$. Provided that the family of triangulations is* regular$^+$*, one has:*

$$\|B_\mu - B_\mu^h\|_{\mathcal{L}(H^\mu(\mathcal{R}))} \lesssim h^{\theta\mu}. \tag{6.3.5}$$

Proof. Using the triangular inequality, [33, Theorem 2.3] and the inverse inequality (6.3.4), it holds:

$$\begin{array}{rcl} \|B_\mu f - B_\mu^h f\|_{H^\mu(\mathcal{R})} & \leq & \|B_\mu f - \pi^0(B_\mu f)\|_{H^\mu(\mathcal{R})} + \|\pi^0(B_\mu f) - B_\mu^h f\|_{H^\mu(\mathcal{R})}, \\ & \lesssim & h^{1-\mu}\|\phi\|_{\mathcal{P}H^1(\mathcal{R})} + h^{\theta-2\mu}\|\pi^0(B_\mu f) - B_\mu^h f\|_{L^2(\mathcal{R})}. \end{array}$$

Using the triangular inequality and estimate (6.2.15) and inequality (6.3.2), we have:

$$\|\pi^0(B_\mu f) - B_\mu^h f\|_{L^2(\mathcal{R})} \leq \|\pi^0(B_\mu f) - B_\mu f\|_{L^2(\mathcal{R})} + \|B_\mu f - B_\mu^h f\|_{L^2(\mathcal{R})} \leq h\,\|f\|_{H^\mu(\mathcal{R})} + h^{2\mu}\,\|f\|_{H^\mu(\mathcal{R})}.$$

We obtain (6.3.5) gathering the above estimates. ◇

Thanks to [206], convergence of the discrete eigenvalues to the exact ones is guaranteed, and so is the absence of spectral pollution:

- Given any closed, non-empty disk $D \subset \mathbb{C}$ such that $D \cap \sigma(B_\mu) = \emptyset$, there exists $h_0 > 0$ such that, for all $h < h_0$, $D \cap \sigma(B_\mu^h) = \emptyset$.
- Given any closed, non-empty disk $D \subset \mathbb{C}$ such that $D \cap \sigma(B_\mu) = \{\lambda\}$, with λ of multiplicity m_λ, there exists $h_0 > 0$ such that, for all $h < h_0$, $D \cap \sigma(B_\mu^h)$ contains exactly m_λ discrete eigenvalues.

Optimal convergence rate

Let the assumptions of Theorem 6.3.1 hold. We determine now the rate of convergence of the eigenvalues in the spirit of [53]. Let $\nu = \lambda^{-1}$ be an eigenvalue of B_μ. For simplicity, let us assume that ν is a simple eigenvalue, and denote by W the associated eigenspace. We remark that $\varphi \mapsto \|\varphi\|_W = \|(\underline{\nu\Sigma}_f)^{\frac{1}{2}}\varphi\|_{L^2(\mathcal{R})}$ is a norm over W([2]), and this norm is induced by the inner product

$$(\varphi, \varphi')_W = (\underline{\nu\Sigma}_f\varphi, \varphi')_{L^2(\mathcal{R})}.$$

According to the absence of spectral pollution, for h small enough, the closest discrete eigenvalue, denoted by ν_h, is also simple ; we denote by W_h the associated eigenspace.

[2] If $\|\varphi\|_W = 0$, then $\underline{\nu\Sigma}_f\varphi = 0$. By definition of W, φ is solution to (6.3.1) with zero right-hand side. Thus, by uniqueness of the solution it follows that $\varphi = 0$.

Definition 6.3.2. *Let $\omega_\nu > 0$ be the regularity exponent of the eigenfunction, i.e. either $W \subset \mathcal{P}H^{1+s}(\mathcal{R})$ for $s < \omega_\nu$ and $W \not\subset \mathcal{P}H^{1+\omega_\nu}(\mathcal{R})$, or $W \subset \mathcal{P}H^{1+\omega_\nu}(\mathcal{R})$ and $W \not\subset \mathcal{P}H^{1+s}(\mathcal{R})$ for $s > \omega_\nu$. Let $\omega = \min(\omega_\nu, m+1)$, where $m \geq 0$ is the order of the RTN finite element.*

Clearly, ω_ν, and as a consequence ω, can be greater than $r_{\max}$. We shall prove that the approximation converges with a rate equal to twice the exponent ω defined above: this result is stated in corollary 6.3.1 at the end of the subsection.
Let $\mu \in [0, r_{\max}[$ be given. As we defined B_μ (resp. B_μ^h), we define A_μ (resp. A_μ^h): for $f \in H^\mu(\mathcal{R})$, we call $A_\mu f = \mathbf{p} \in \mathbf{H}(\mathrm{div}, \mathcal{R})$ (resp. $A_\mu^h f = \mathbf{p}_h \in \mathbf{Q}_h^k$) the first component of the couple $(\mathbf{p}, \phi)$ (resp. $(\mathbf{p}_h, \phi_h)$) that solves (6.2.4) (resp. (6.2.11)) with source $S_f = \underline{\nu\Sigma}_f f$.

Proposition 6.3.1 (`[A2, Proposition 4.17]`). *Let ω be as in Definition 6.3.2. For every φ in W, the following inequalities hold:*

$$\|(B_\mu - B_\mu^h)\varphi\|_{L^2(\mathcal{R})} \lesssim h^\omega \|\varphi\|_W \quad \text{and} \quad \|(A_\mu - A_\mu^h)\varphi\|_{\mathbf{H}(\mathrm{div},\mathcal{R})} \lesssim h^\omega \|\varphi\|_W .$$

Proof. These two inequalities come from the first Strang's Lemma. The method is the same as for Theorem 6.2.7 (see remark 6.2.2 for the "smooth" case). Here, we use the equivalence of all norms on W to state the result. ◇

Introducing $\delta(Z, Z') = \sup_{z \in Z,\, \|z\|_{L^2(\mathcal{R})}=1} \inf_{z' \in Z'} \|z - z'\|_{L^2(\mathcal{R})}$ for Z, Z' closed subspaces of $L^2(\mathcal{R})$, the gap between W and W_h is defined by: $\hat{\delta}(W, W_h) = \max[\delta(W, W_h), \delta(W_h, W)]$. It allows us to evaluate the approximation of the continuous eigenfunctions by their discrete counterparts. Classically, this gap can be bounded with the help of Proposition 6.3.1, following [206, Theorem 1]:

$$\hat{\delta}(W, W_h) \lesssim h^\omega. \tag{6.3.6}$$

Let us now define E_h as the projector from $L^2(\mathcal{R})$ onto W_h such that

$$\forall \varphi \in L^2(\mathcal{R}),\ \forall \psi_h \in W_h,\ (\underline{\nu\Sigma}_f(\varphi - E_h\varphi), \psi_h)_{L^2(\mathcal{R})} = 0. \tag{6.3.7}$$

Lemma 6.3.2 (`[A2, Lemma 4.18]`). *The operators E_h and B_μ^h commute.*

Let F_h be the restriction of E_h to W. One has the following simple results as a consequence of the gap property.

Lemma 6.3.3 (`[A2, Lemma 4.19]`). *For h small enough, F_h is a bijection from W to W_h and*

$$\forall \varphi \in W,\ \left\| (\underline{\nu\Sigma}_f)^{\frac{1}{2}} (\varphi - F_h\varphi) \right\|_{L^2(\mathcal{R})} \lesssim h^\omega \|\varphi\|_W. \tag{6.3.8}$$

Let $\mathcal{S}_h = F_h^{-1} E_h - I \in \mathcal{L}(L^2(\mathcal{R}))$ for h small enough.

Lemma 6.3.4 (`[A2, Lemma 4.20]`). *For h small enough, $W \subset \ker(\mathcal{S}_h)$; the sequence of operators $(\mathcal{S}_h)_h$ is uniformly bounded.*

One can then prove an "orthogonality" result involving $\mathcal{S}_h$.

Proposition 6.3.2 (`[A2, Proposition 4.21]`). *For all f in $L^2(\mathcal{R})$ and φ_h in W_h, one has for h small enough $(\underline{\nu\Sigma}_f \mathcal{S}_h f, \varphi_h)_{L^2(\mathcal{R})} = 0$.*

To obtain an optimal convergence rate, we restrict the operators B_μ and B_μ^h to the eigenspace W. We denote finally by $\hat{B}_\mu$ and $\hat{B}_\mu^h$ the operators, from W to itself, $\hat{B}_\mu = B_\mu|_W$ and $\hat{B}_\mu^h = F_h^{-1} B_\mu^h F_h$.

Theorem 6.3.2 (`[A2, Theorem 4.22]`). *Let ω be as in Definition 6.3.2. For h small enough, the following estimate holds:*

$$\|\hat{B}_\mu - \hat{B}_\mu^h\|_{\mathcal{L}(W)} \lesssim h^{2\omega}. \tag{6.3.9}$$

Proof. Let $\varphi' \in W$. It holds: $(\hat{B}_\mu - \hat{B}^h_\mu)\varphi' = B_\mu\varphi' - F_h^{-1}B^h_\mu F_h\varphi' = B_\mu\varphi' - F_h^{-1}B^h_\mu E_h\varphi'$ by definition. Hence, using successively lemmas 6.3.2 and 6.3.4, and the definition of $\mathcal{S}_h$, it holds (with $\delta B_\mu = B_\mu - B^h_\mu$):

$$(\hat{B}_\mu - \hat{B}^h_\mu)\varphi' = B_\mu\varphi' - F_h^{-1}E_hB^h_\mu\varphi' = \delta B_\mu\varphi' + B^h_\mu\varphi' - F_h^{-1}E_hB^h_\mu\varphi' + \mathcal{S}_hB_\mu\varphi' = \delta B_\mu\varphi' + \mathcal{S}_h\delta B_\mu\varphi'.$$

Given φ, $\varphi' \in W$, we bound $|(\varphi, (\hat{B}_\mu - \hat{B}^h_\mu)\varphi')_W| = |(\underline{\nu\Sigma}_f\varphi, (\hat{B}_\mu - \hat{B}^h_\mu)\varphi')_0|$ by $|(\underline{\nu\Sigma}_f\varphi, \delta B_\mu\varphi')_0| + |(\underline{\nu\Sigma}_f\varphi, \mathcal{S}_h\delta B_\mu\varphi')_0|$, where $(\cdot,\cdot)_0 := (\cdot,\cdot)_{L^2(\mathcal{R})}$. Noticing that $a(\cdot,\cdot)$ and $t(\cdot,\cdot)$ are symmetric, one show that (with $\delta A_\mu = A_\mu - A^h_\mu$):

$$(\underline{\nu\Sigma}_f\varphi, \delta B_\mu\varphi')_0 = a(\delta A_\mu\varphi, \delta A'_\varphi) + b(\delta A_\mu\varphi', \delta B_\mu\varphi) + b(\delta A_\mu\varphi, \delta B_\mu\varphi') + t(\delta B_\mu\varphi, \delta B_\mu\varphi').$$

The first part of the sum can then be bound as follows:

$$\begin{array}{rcl} |(\underline{\nu\Sigma}_f\varphi, \delta B_\mu\varphi')_0| & \lesssim & \|\delta A_\mu\varphi\|_0\|\delta A_\mu\varphi'\|_0 + \|\operatorname{div}\delta A_\mu\varphi'\|_0\|\delta B_\mu\varphi\|_0 + \|\operatorname{div}\delta A_\mu\varphi\|_0\|\delta B_\mu\varphi'\|_0 + \|\delta B_\mu\varphi\|_0\|\delta B_\mu\varphi'\|_0 \\ & \lesssim & h^{2\omega}\|\varphi\|_W\|\varphi'\|_W. \end{array}$$

Using successively Proposition 6.3.2 with $f = \delta B_\mu\varphi'$ and $\varphi_h = F_h\varphi$; the uniform continuity of $\mathcal{S}_h$ in h; the first inequality of Proposition 6.3.1 and the estimation (6.3.8), we bound the second part as follows (with $\delta\varphi =: \varphi - F_h\varphi$):

$$|(\underline{\nu\Sigma}_f\varphi, \mathcal{S}_h\delta B_\mu\varphi')_0| = |(\underline{\nu\Sigma}_f\delta\varphi, \mathcal{S}_h\delta B_\mu\varphi')_0| \le \|\underline{\nu\Sigma}_f\delta\varphi\|_0\|\mathcal{S}_h\delta B_\mu\varphi'\|_0 \lesssim \|\underline{\nu\Sigma}_f\delta\varphi\|_0\|\delta B_\mu\varphi'\|_0 \lesssim h^{2\omega}\|\varphi\|_W\|\varphi'\|_W.$$

Therefore we have obtained (6.3.9). ◇

From this estimation and the work of Osborn in [206, Theorem 2], one derives an optimal estimate on the error on the eigenvalues.

Corollary 6.3.1. *Let ω be as in Definition 6.3.2. Then for h small enough, the error on the eigenvalue is given by $|\nu - \nu_h| \lesssim h^{2\omega}$.*

Remark 6.3.1. *If ν has an algebraic multiplicity $m_\nu > 1$, the previous analysis and the a priori estimate are still valid with $\nu_h = \frac{1}{m_\nu}\sum_{i=1}^{m_\nu}\nu_{h,i}$, where $(\nu_{h,i})_{i=1,m_\nu}$ are the m discrete eigenvalues closest to ν, see again [206, Theorem 2].*

6.4 Domain decomposition

Studying numerically the steady state of a nuclear core reactor is expensive, in terms of memory storage and computational time. In order to address both requirements, one can use a domain decomposition method, implemented on a parallel computer. We present here such two methods for the mixed neutron diffusion equations, discretized with Raviart-Thomas-Nédélec finite elements. The first method is based on the Schwarz iterative algorithm with Robin interface conditions to handle communications, see Section 6.5. In the second method, communications are handled with Lagrange multipliers, so that the domain decomposition method can be non matching in the sense that the traces of the finite element spaces may not fit at the interface between the subdomains, see Section 6.6. We analyse both method from the continuous point of view to the discrete point of view, and we give some numerical results in a realistic highly heterogeneous $3D$ configuration. We refer to [212, 134] for an overview of domain decomposition methods for partial differential equations. See [218] for a domain decomposition method for crack problems.

Let $\mathcal{R}$ be a domain of $\mathbb{R}^d$. We consider the following problem:

$$\text{Find } \phi \in H^1_0(\mathcal{R}) \text{ such that} \qquad -\operatorname{div}(D\,\nabla\phi) + \Sigma_a\,\phi = S_f \text{ on } \mathcal{R}. \tag{6.4.1}$$

where $S_f \in L^2(\mathcal{R})$, D, $\Sigma_a \in L^\infty(\mathcal{R})$ satisfying assumptions given in 6.1.7.

Problem (6.4.1) can also be written as a mixed problem with two unknowns $(\mathbf{p}, \phi)$:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{p},\phi) \in \mathbf{H}(\operatorname{div},\mathcal{R}) \times H^1_0(\mathcal{R}) \text{ such that} \\ \operatorname{div}\mathbf{p} + \Sigma_a\,\phi = S_f \text{ and } D^{-1}\mathbf{p} + \nabla\phi = 0 \text{ in } \mathcal{R}. \end{array}\right. \tag{6.4.2}$$

Wellposedness of problems (6.4.1) and (6.4.2) are stated in Theorems 6.2.1 and 6.2.2. Consider a partition

of $\mathcal{R}$ in N (non empty) subdomains $\{\widetilde{\mathcal{R}}_i\}_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}$[3], with $\tilde{\mathcal{I}}_{\mathcal{R}} := \{1,\dots,\tilde{N}_{\mathcal{R}}\}$: $\overline{\widetilde{\mathcal{R}}} = \cup_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} \overline{\widetilde{\mathcal{R}}_i}$. The partition is such that for all $(i,j) \in \tilde{\mathcal{I}}_{\mathcal{R}} \times \tilde{\mathcal{I}}_{\mathcal{R}}$, $i \neq j$, $\widetilde{\mathcal{R}}_i \cap \widetilde{\mathcal{R}}_j = \emptyset$. For $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, we let $\mathbf{n}_i$ be the unit outgoing normal to $\partial\widetilde{\mathcal{R}}_i$ and, for $p \in [1,\infty]$ and $z \in L^p(\mathcal{R})$, we let here $z_i := z_{|\widetilde{\mathcal{R}}_i}$.

We set for $i \neq j$, $\Gamma_{ij} = int(\overline{\widetilde{\mathcal{R}}_i} \cap \overline{\widetilde{\mathcal{R}}_j})$ if the Hausdorff dimension of $\overline{\widetilde{\mathcal{R}}_i} \cap \overline{\widetilde{\mathcal{R}}_j}$ is equal to $d-1$ (the subdomains are neighbours); and $\Gamma_{ij} = \emptyset$ otherwise. We have by construction: $\Gamma_{ij} = \Gamma_{ji}$, we will use indifferently both notations. Remark that $\mathbf{n}_{i|\Gamma_{ij}} = -\mathbf{n}_{j|\Gamma_{ij}}$. On the interface Γ_{ij}, we set $\mathbf{n}_{ij} = \mathbf{n}_{\min(i,j)|\Gamma_{ij}}$. We may write: $\partial_{\mathbf{n}_i}\phi = \nabla\phi\cdot\mathbf{n}_i$ on $\partial\widetilde{\mathcal{R}}_i$ and $\partial_{\mathbf{n}_{ij}}\phi = \nabla\phi\cdot\mathbf{n}_{ij}$ on Γ_{ij}.
We then introduce the following index and pair spaces:

$$\begin{array}{rcll} \mathcal{I}_S & := & \{i \in \tilde{\mathcal{I}}_{\mathcal{R}} \,|\, \partial\widetilde{\mathcal{R}}_i \cap \Gamma_S = \partial\widetilde{\mathcal{R}}_i\}, & \text{(inner subdomains)} \\ \mathcal{I}_{\widetilde{\mathcal{R}}_i} & := & \{j \in \tilde{\mathcal{I}}_{\mathcal{R}} \,|\, \Gamma_{ij} \neq \emptyset\}, \quad \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}} & \text{(subdomains neighbouring } \widetilde{\mathcal{R}}_i) \\ \mathcal{IJ} & := & \{(i,j) \in \tilde{\mathcal{I}}_{\mathcal{R}} \times \tilde{\mathcal{I}}_{\mathcal{R}},\, i<j \,|\, \Gamma_{ij} \neq \emptyset\}, & \text{(neighbouring subdomains pairs).} \end{array} \tag{6.4.3}$$

For $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, we call Σ_i the union of the interfaces of $\partial\widetilde{\mathcal{R}}_i$: $\Sigma_i = \cup_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \Gamma_{ij}$. We define the interface Γ_S and the wirebasket $\partial\Gamma_W$ by:

$$\Gamma_S = \bigcup_{(i,j)\in\mathcal{IJ}} \overline{\Gamma_{ij}}, \quad \partial\Gamma_W = \bigcup_{(i,j)\in\mathcal{IJ}} \partial\Gamma_{ij}.$$

The resulting interface Γ_S doen't need to coincide with the physical interface between the cells. When $d=2$, the wirebasket is made up of the vertices of the interface, which are isolated points.
When $d=3$, the wirebasket is made up of the vertices and (open) edges of the interface.
Wirebasket points not located on the boundary $\partial\widetilde{\mathcal{R}}$ are crossing points.
For $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$ we introduce the open $\Gamma_i = \partial\widetilde{\mathcal{R}}_i \backslash \overline{\Gamma_S}$. By construction, $\partial\widetilde{\mathcal{R}}_i = \Sigma_i \cup (\partial\Gamma_W \cap \partial\widetilde{\mathcal{R}}_i) \cup \Gamma_i$, and the partition is disjoint.

We define for all $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$ the subspace of functions of $H^1(\widetilde{\mathcal{R}}_i)$ vanishing on $\partial\mathcal{R}$:

$$V_i := \{\psi \in H^1(\widetilde{\mathcal{R}}_i) \,|\, \psi_{|\Gamma_i} = 0\}. \tag{6.4.4}$$

We introduce the following product spaces endowed with their product norms:

$$\begin{array}{rcll} \widetilde{\mathcal{P}}V & = & \displaystyle\prod_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} V_i, & \|\psi\|_{\widetilde{\mathcal{P}}V} = \left(\displaystyle\sum_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} \|\psi_i\|^2_{H^1(\widetilde{\mathcal{R}}_i)}\right)^{1/2}, \\ \widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div},\mathcal{R}) & = & \displaystyle\prod_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} \mathbf{H}(\mathrm{div},\widetilde{\mathcal{R}}_i), & \|\mathbf{q}\|_{\widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div},\mathcal{R})} = \left(\displaystyle\sum_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} \|\mathbf{q}_i\|^2_{\mathbf{H}(\mathrm{div},\widetilde{\mathcal{R}}_i)}\right)^{1/2}. \end{array} \tag{6.4.5}$$

Remark 6.4.1. *These spaces can be described equivalently as follows:*

$$\widetilde{\mathcal{P}}V = \{\psi \in L^2(\mathcal{R}) \,|\, \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}},\, \psi_i \in V_i\}, \quad \widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div},\mathcal{R}) = \{\mathbf{q} \in \mathbf{L}^2(\mathcal{R}) \,|\, \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}},\, \mathbf{q}_i \in \mathbf{H}(\mathrm{div},\widetilde{\mathcal{R}}_i)\}.$$

For $\psi \in \widetilde{\mathcal{P}}V$, we introduce the jump of ψ on Γ_{ij}:

$$\forall (i,j) \in \mathcal{IJ}, \quad [\psi]_{ij} := \psi_{i|\Gamma_{ij}} - \psi_{j|\Gamma_{ij}}. \tag{6.4.6}$$

The jump $[\psi]_{ij}$ is well defined in $H^{\frac{1}{2}}(\Gamma_{ij})$. The global jump $[\psi]$ of ψ on Γ_S is defined by:

$$\forall (i,j) \in \mathcal{IJ}, \quad [\psi]_{|\Gamma_{ij}} := [\psi]_{ij}\,. \tag{6.4.7}$$

For all pair $(i,j) \in \mathcal{IJ}$, we denote by $H_{ij} := \widetilde{H}^{\frac{1}{2}}(\Gamma_{ij})$, the space of $H^{\frac{1}{2}}(\Gamma_{ij})$ functions which extension

[3]Each sub-domain has a sufficiently regular boundary.

by zero in $\partial\widetilde{\mathcal{R}}_i$ is in $H^{\frac{1}{2}}(\partial\widetilde{\mathcal{R}}_i)$ (cf. §4.1). We can prove that $H_{ij} = H_{ji}$ (cf. e.g. [16, chap. 2]). We call H'_{ij} the dual space of H_{ij} (see e.g. [151]).
For all $\mathbf{q} \in \widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div}, \mathcal{R})$, we introduce the jump of its normal component on Γ_{ij}:

$$\forall (i,j) \in \mathcal{IJ}, \quad [\mathbf{q}\cdot\mathbf{n}]_{ij} := \mathbf{q}_i \cdot \mathbf{n}_{i|\Gamma_{ij}} + \mathbf{q}_j \cdot \mathbf{n}_{j|\Gamma_{ij}}. \tag{6.4.8}$$

The jump $[\mathbf{q}\cdot\mathbf{n}]_{ij}$ is well defined in H'_{ij}, cf. §4.2.2. The *global jump* $[\mathbf{q}\cdot\mathbf{n}]$ of the normal component on Γ_S is defined by:

$$\forall (i,j) \in \mathcal{IJ}, \quad [\mathbf{q}\cdot\mathbf{n}]_{|\Gamma_{ij}} := [\mathbf{q}\cdot\mathbf{n}]_{ij}\,. \tag{6.4.9}$$

By definition, it holds: $[\mathbf{q}\cdot\mathbf{n}] \in \widetilde{M}$, where $\widetilde{M}$ is the product space: $\widetilde{M} := \prod_{(i,j)\in\mathcal{IJ}} H'_{ij}$. Finally, we introduce the following Hilbert spaces:

$$\begin{aligned} M &= \left\{ \psi_S \in \prod_{(i,j)\in\mathcal{IJ}} L^2(\Gamma_{ij}) \right\}, \quad \|\psi_S\|_M = \left(\sum_{(i,j)\in\mathcal{IJ}} \|\psi_S\|^2_{L^2(\Gamma_{ij})} \right)^{1/2}; \\ \widetilde{\mathbf{Q}} &= \left\{ \mathbf{q} \in \widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div}, \mathcal{R}) \,|\, [\mathbf{q}\cdot\mathbf{n}] \in M \right\}, \quad \|\mathbf{q}\|_{\widetilde{\mathbf{Q}}} := \left(\|\mathbf{q}\|^2_{\widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div},\mathcal{R})} + \|[\mathbf{q}\cdot\mathbf{n}]\|^2_M \right)^{1/2}. \end{aligned} \tag{6.4.10}$$

By construction, $M \subset \widetilde{M}$. In the definition of elements $\mathbf{q} = (\mathbf{q}_i)_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}$ of $\widetilde{\mathbf{Q}}$, it is important to note that for $(i,j) \in \mathcal{IJ}$, *it is not required that* $\mathbf{q}_i \cdot \mathbf{n}_{i|\Gamma_{ij}}, \mathbf{q}_j \cdot \mathbf{n}_{j|\Gamma_{ij}} \in L^2(\Gamma_{ij})$. This definition is based on the fact that, as far as problem (6.4.1) is concerned, the normal trace of the gradient on the interface, $\nabla\phi\cdot\mathbf{n}_{|\Gamma_S}$, may not belong to M. On the other hand, since the jump $[D\,\nabla\phi\cdot\mathbf{n}]$ is zero, it systematically belongs to M. We'll use this observation to build a conformal variational formulation in $\widetilde{\mathbf{Q}}$.

The following formulas for integration by parts will be useful for constructing variational formulations, following the work of [A6]. According to the integration by parts formula (4.2.2), we have for all $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$:

$$\forall (\mathbf{q}_i, \psi_i) \in \mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i) \times H^1(\widetilde{\mathcal{R}}_i), \quad \int_{\widetilde{\mathcal{R}}_i} (\psi_i \ \mathrm{div}\,\mathbf{q}_i + \nabla\psi_i \cdot \mathbf{q}_i) = \langle \mathbf{q}_i\cdot\mathbf{n}_i, \psi_i \rangle_{H^{1/2}(\partial\widetilde{\mathcal{R}}_i)}.$$

We introduce, for $H \in \{H^1(\mathcal{R}), H^1_0(\mathcal{R}), (V_i)_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}, \widetilde{\mathcal{P}}V\}$:

$$H_- := \{\psi \in H \,|\, \psi = 0 \text{ in a neighbourhood of } \partial\Gamma_W\}.$$

We now recall some density results proven in [119].

Theorem 6.4.1. *Let $H \in \{H^1(\mathcal{R}), H^1_0(\mathcal{R}), (V_i)_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}, \widetilde{\mathcal{P}}V\}$: H_- is dense in H.*

Using one of these density results, we can establish a new formula for integration by parts.

Corollary 6.4.1. *We have the integration by parts formula:*

$$\forall (\mathbf{q}, \psi) \in \widetilde{\mathbf{Q}} \times H^1_0(\mathcal{R}), \quad \int_{\mathcal{R}} (\nabla\psi\cdot\mathbf{q} + \psi \ \mathrm{div}\,\mathbf{q}) = \int_{\Gamma_S} [\mathbf{q}\cdot\mathbf{n}]\ \psi_{|\Gamma_S}. \tag{6.4.11}$$

Proof. For $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$ the IBP formula (4.2.2) (with duality brackets cut out) in $\widetilde{\mathcal{R}}_i$ becomes, using the H_{ij} trace space:

$$\forall (\mathbf{q}_i, \psi_i) \in \mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i) \times (V_i)_-\,, \quad \int_{\widetilde{\mathcal{R}}_i} (\nabla\psi_i\cdot\mathbf{q}_i + \psi_i \ \mathrm{div}\,\mathbf{q}_i) = \sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \langle \mathbf{q}_i\cdot\mathbf{n}_i, \psi_i\rangle_{H_{ij}}.$$

Consider $\mathbf{q} \in \widetilde{\mathbf{Q}}$ and $\psi \in \left(H^1_0(\mathcal{R})\right)_-$. For all $(i,j) \in \mathcal{IJ}$, we have $\psi_{i|\Gamma_{ij}} = \psi_{j|\Gamma_{ij}} = \psi_{|\Gamma_{ij}}$. In particular, we can define

$\mu = \psi_{|\Gamma_S} \in M$, and we can write:[4]

$$\begin{aligned}\int_{\mathcal{R}} (\nabla\,\psi\cdot\mathbf{q} + \psi\ \mathrm{div}\,\mathbf{q}) &= \sum_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} \int_{\widetilde{\mathcal{R}}_i} (\nabla\,\psi_i\cdot\mathbf{q}_i + \psi_i\ \mathrm{div}\,\mathbf{q}_i) + \sum_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}\sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \langle \mathbf{q}_i\cdot\mathbf{n}_i, \psi_i\rangle_{H_{ij}} \\ &= \sum_{(i,j)\in\mathcal{IJ}} \langle \mathbf{q}_i\cdot\mathbf{n}_i + \mathbf{q}_j\cdot\mathbf{n}_j, \psi_{|\Gamma_{ij}}\rangle_{H_{ij}} = \sum_{(i,j)\in\mathcal{IJ}} \langle [\mathbf{q}\cdot\mathbf{n}]_{ij}\,, \mu\rangle_{H_{ij}} = \int_{\Gamma_S} [\mathbf{q}\cdot\mathbf{n}]\ \mu.\end{aligned}$$

According to Theorem 6.4.1, the result is still true for $\psi \in H_0^1(\mathcal{R})$ and $\mathbf{q} \in \widetilde{\mathbf{Q}}$. ◇

We now consider problem (6.4.2) posed into local problems.

Theorem 6.4.2 ([A6, Theorem 1]). *Problem* (6.4.2) *is equivalent to the following problem:*

$$\left\{\begin{array}{lrcll} \text{Find } (\mathbf{p},\phi) \in \widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div},\mathcal{R}) \times \widetilde{\mathcal{P}}V \text{ such that:} & & & & \\ \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}}, & \mathrm{div}\,\mathbf{p}_i + \Sigma_{a,i}\,\phi_i &=& S_{f,i} & \text{in } \widetilde{\mathcal{R}}_i \quad (0) \\ \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}}, & (D_i)^{-1}\,\mathbf{p}_i + \nabla\,\phi_i &=& 0 & \text{in } \widetilde{\mathcal{R}}_i \quad (\mathrm{i}) \\ \forall (i,j) \in \mathcal{IJ}, & [\phi]_{ij} &=& 0 & \text{on } \Gamma_{ij} \quad (\mathrm{ii}) \\ \forall (i,j) \in \mathcal{IJ}, & [\mathbf{p}\cdot\mathbf{n}]_{ij} &=& 0 & \text{on } \Gamma_{ij} \quad (\mathrm{iii}) \end{array}\right. , \tag{6.4.12}$$

with the identification for all $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, $(\mathbf{p}_i,\phi_i) = (\mathbf{p}_{|\widetilde{\mathcal{R}}_i}, \phi_{|\widetilde{\mathcal{R}}_i})$ *and* $S_{f,i} = S_{f|\widetilde{\mathcal{R}}_i}$.

We can write (6.4.12)-(ii) and (6.4.12)-(iii) this way:

$$\forall (i,j) \in \mathcal{IJ} \left\{\begin{array}{rcll} \phi_i &=& \phi_j & \text{on } \Gamma_{ij} \quad (\mathrm{i}) \\ \mathbf{p}_i\cdot\mathbf{n}_{ij} &=& \mathbf{p}_j\cdot\mathbf{n}_{ij} & \text{on } \Gamma_{ij} \quad (\mathrm{ii}) \end{array}\right. . \tag{6.4.13}$$

Notice that the trace $\phi_{i|\Gamma_{ij}} \in H_{ij}$, and can be extended by zero on $\partial\widetilde{\mathcal{R}}_j$ to obtain a function of $H^{\frac{1}{2}}(\partial\widetilde{\mathcal{R}}_j)$. By virtue of the surjectivity of the trace application (see Theorem 4.2.1), there is a lifting in V_j, which by definition belongs to the space V_j^i such that:

$$V_j^i := \{v \in V_j \,|\, v_{|\partial\widetilde{\mathcal{R}}_j\setminus\Gamma_{ij}} = 0\}. \tag{6.4.14}$$

Next, we will see study different domain decomposition methods: the Optimized Schwarz Method in Section 6.5 and the DD+L^2-jumps method in Section 6.6. To perform industrial computations, we use some components of the APOLLO3®, the latest neutronics code of CEA/DES, developed in collaboration with EDF and Orano [164, P8, P10]. In particular, we use the MINOS solver, one of the deterministic solvers, to compute numerically k_{eff}. More specifically, the MINOS solver [28, 29, P8] is a $3D$, multigroup SP_N solver on structured cartesian and hexagonal grids. The unknown $(\mathbf{p},\phi)$ is discretized using $\mathrm{RT}_{[k]}$ finites elements on structured cartesian grids. Note that the DD version is parallelized in the APOLLO3® code, contrary to the plain version. Hence, computational times are greatly reduced: we refer to [A11] for the analyses of algorithms and their parallelization.

6.5 Optimized Schwarz method [A11, A10, P6]

Schwarz iterative algorithm was designed in 1869 [229] in order to study the Laplace operator on irregular domains: Hermann Schwarz's original idea was to split the original domain into overlapping regular subdomains (such as rectangles or disks) and to write an iterative solver using solutions on the regular subdomains to converge to the global solution on the irregular domain. Convergence was proved by Lions in 1988 [195], we proposed in 1990 to modify the interface conditions (6.4.13) in order to achieve convergence without overlap [196, 212]. Optimized Schwarz method is widely used for other models: Maxwell's equation [108, 133], Helmholtz equation [219, 106], Stokes problem [78], acoustics [107], parabolic problems [77].

[4] Strictly speaking, the integral on $\mathcal{R}$ is valid for extensions of $\mathbf{q}$ and $(\mathrm{div}\,\mathbf{q})$ in $\mathcal{R}$.

Suppose here that the solution $(\mathbf{p}, \phi)$ is such that for all $(i,j) \in \tilde{\mathcal{I}}_{\mathcal{R}} \times \mathcal{I}_{\widetilde{\mathcal{R}}_i}$, $\mathbf{p}_i \cdot \mathbf{n}_{ij|\Gamma_{ij}} \in L^2(\Gamma_{ij})$, and introduce spaces $\mathbf{Q}_i$, for $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$ endowed with the inner norm:

$$\mathbf{Q}_i := \left\{ \mathbf{q} \in \mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i) \,:\, \mathbf{q} \cdot \mathbf{n}_{i|\Sigma_i} \in L^2(\Sigma_i) \right\}, \ \|\mathbf{q}\|^2_{\mathbf{Q}_i} := \|\mathbf{q}\|^2_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)} + \|\mathbf{q}\|^2_{L^2(\Sigma_i)}. \tag{6.5.1}$$

For all $(i,j) \in \tilde{\mathcal{I}}_{\mathcal{R}} \times \mathcal{I}_{\widetilde{\mathcal{R}}_i}$, we introduce couples $(\alpha^i_{ij}, \alpha^j_{ij})$ of positive coefficients such that $\alpha^i_{ij} + \alpha^j_{ij} > 0$. The Dirichlet (resp. Neumann) interface conditions (6.4.13)-(ii) (resp. (6.4.13)-(iii)) can be combined to write the following interface conditions:

$$\text{For all } (i,j) \in \mathcal{IJ}, \left\{ \begin{array}{rcll} \mathbf{p}_i \cdot \mathbf{n}_{ij} + \alpha^i_{ij} \phi_i & = & \mathbf{p}_j \cdot \mathbf{n}_{ij} + \alpha^i_{ij} \phi_j & \text{on } \Gamma_{ij}, \\ \mathbf{p}_j \cdot \mathbf{n}_{ij} - \alpha^j_{ij} \phi_j & = & \mathbf{p}_i \cdot \mathbf{n}_{ij} - \alpha^j_{ij} \phi_i & \text{on } \Gamma_{ij}. \end{array} \right. \tag{6.5.2}$$

They can also be expressed as a single equation as follows:

$$\text{For all } (i,j) \in \mathcal{I} \times \mathcal{I}_{\widetilde{\mathcal{R}}_i}, \quad \mathbf{p}_i \cdot \mathbf{n}_i + \alpha^i_{ij} \phi_i = -\mathbf{p}_j \cdot \mathbf{n}_j + \alpha^i_{ij} \phi_j \text{ on } \Gamma_{ij}. \tag{6.5.3}$$

Problem (6.4.12) is equivalent to the following problem:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{p}, \phi) \in \prod_{i \in \tilde{\mathcal{I}}_{\mathcal{R}}} \mathbf{Q}_i \times \widetilde{\mathcal{P}} V \text{ such that for all } i \in \tilde{\mathcal{I}}_{\mathcal{R}} \text{ and for all } j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}: \\ \begin{array}{rcll} \mathrm{div}\, \mathbf{p}_i + \Sigma_{a,i}\, \phi_i & = & S_{f,i} & \text{in } \widetilde{\mathcal{R}}_i \quad (0) \\ (D_i)^{-1}\, \mathbf{p}_i + \nabla\, \phi_i & = & 0 & \text{in } \widetilde{\mathcal{R}}_i \quad \text{(i)} \\ \mathbf{p}_i \cdot \mathbf{n}_i + \alpha^i_{ij} \phi_i & = & -\mathbf{p}_j \cdot \mathbf{n}_j + \alpha^i_{ij} \phi_j & \text{on } \Gamma_{ij} \quad \text{(ii)} \end{array} \end{array} \right. . \tag{6.5.4}$$

As we don't know the values of $\phi_{i|\Sigma}$ nor $\mathbf{p}_i \cdot \mathbf{n}_{i|\Sigma}$, we propose to solve problems (6.5.4) iteratively, locally in subdomains $(\widetilde{\mathcal{R}}_i)_{i \in \tilde{\mathcal{I}}_{\mathcal{R}}}$, and to approximate the interface conditions. In order to solve Problem (6.5.4), we now propose the following iterative Schwarz algorithm:

$$\left\| \begin{array}{l} \textbf{initialization} \\ \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}}, \quad (\mathbf{p}^0_i, \phi^0_i) \in \mathbf{Q}_i \times V_i \textbf{ is given.} \\ \textbf{iterations: for } \ell = 1, \ldots, \quad \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}}, \\ \\ \left\{ \begin{array}{l} \text{Find } (\mathbf{p}^\ell_i, \phi^\ell_i) \in \mathbf{Q}_i \times V_i \text{ such that} \\ \begin{array}{rcll} \mathrm{div}\, \mathbf{p}^\ell_i + \Sigma_{a,i}\, \phi^\ell_i & = & S_{f,i} & \text{in } \widetilde{\mathcal{R}}_i \quad (0) \\ (D_i)^{-1}\, \mathbf{p}^\ell_i + \nabla\, \phi^\ell_i & = & 0 & \text{in } \widetilde{\mathcal{R}}_i \quad \text{(i)} \\ \forall j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}, \ -\mathbf{p}^\ell_i \cdot \mathbf{n}_i + \alpha^i_{ij} \phi^\ell_i & = & \mathbf{p}^{\ell-1}_j \cdot \mathbf{n}_j + \alpha^i_{ij} \phi^{\ell-1}_j & \text{on } \Gamma_{ij} \quad \text{(ii)} \end{array} \end{array} \right. , \\ \\ \textbf{untill convergence.} \end{array} \right. \tag{6.5.5}$$

The initial choice is assumed to be such that for all $(i,j) \in \tilde{\mathcal{I}}_{\mathcal{R}} \times \mathcal{I}_{\widetilde{\mathcal{R}}_i}$, $\mathbf{p}^0_i \cdot \mathbf{n}_{ij|\Gamma_{ij}} \in L^2(\Gamma_{ij})$.
We can prove (cf. §A.2) the

Theorem 6.5.1 ([B1, Theorem 11.1]). *Suppose that for all $(i,j) \in \mathcal{IJ}$, $\alpha^i_{ij} = \alpha^j_{ij}$. Then the sequence $((\mathbf{p}^\ell_i, \phi^\ell_i))_{i \in \tilde{\mathcal{I}}_{\mathcal{R}}})_\ell$ converges to the solution to (6.4.2) in $\widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div}, \mathcal{R}) \times \widetilde{\mathcal{P}} V$.*

Furthermore, the asymptotic study of Nataf and Nier [201] makes it possible to optimise the choice of parameters α^i_{ij} [A11, §5.5]. At the lowest order, the optimal parameters are such that:

$$\alpha^i_{ij} = (\Sigma_{a,j}\, D_j)^{\frac{1}{2}}. \tag{6.5.6}$$

We refer to [84] for a rigorous study on the optimized parameters.

6.5.1 Variational formulation

To build the variational formulation associated to Problem (6.5.5), we first multiply (6.5.5)-(i) by $\mathbf{q}_i \in \mathbf{Q}_i$ an we integrate over $\widetilde{\mathcal{R}}_i$, using integration by parts (4.2.2):

$$\int_{\widetilde{\mathcal{R}}_i} \left((D_i)^{-1}\mathbf{p}_i^\ell \cdot \mathbf{q}_i - \phi_i^\ell \,\mathrm{div}\,\mathbf{q}_i\right) + \sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} \phi_i^\ell \,\mathbf{q}_i\cdot\mathbf{n}_i = 0. \tag{6.5.7}$$

According to the transmission condition (6.5.5)-(ii): $\phi_i^\ell = (\alpha_{ij}^i)^{-1}\mathbf{p}_i^\ell\cdot\mathbf{n}_i + (\alpha_{ij}^i)^{-1}\mathbf{p}_j^{\ell-1}\cdot\mathbf{n}_j + \phi_j^{\ell-1}$, so that substituting ϕ_i^ℓ in Eq. (6.5.7), it reads:

$$\begin{aligned}&\int_{\widetilde{\mathcal{R}}_i} \left((D_i)^{-1}\mathbf{p}_i^\ell \cdot \mathbf{q}_i - \phi_i^\ell \,\mathrm{div}\,\mathbf{q}_i\right) + \sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} (\alpha_{ij}^i)^{-1}\mathbf{p}_i^\ell\cdot\mathbf{n}_i\,\mathbf{q}_i\cdot\mathbf{n}_i\\ &= -\sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} (\alpha_{ij}^i)^{-1}\mathbf{p}_j^{\ell-1}\cdot\mathbf{n}_j\,\mathbf{q}_i\cdot\mathbf{n}_i - \sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} \phi_j^{\ell-1}\,\mathbf{q}_i\cdot\mathbf{n}_i.\end{aligned} \tag{6.5.8}$$

We consider the product spaces $\mathbf{X}_i^+ = \mathbf{Q}_i \times L^2(\widetilde{\mathcal{R}}_i)$ for $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, endowed with the product norm

$$\|(\mathbf{q},\psi)\|_{\mathbf{X}_i^+} = \left(\|\mathbf{q}\|^2_{\mathbf{H}(\mathrm{div},\widetilde{\mathcal{R}}_i)} + \|\mathbf{q}\cdot\mathbf{n}_i\|^2_{L^2(\Sigma_i)} + \|\psi\|^2_{L^2(\widetilde{\mathcal{R}}_i)}\right)^{1/2}.$$

In this case, we cannot keep the trace of the $\phi_j^{\ell-1}$ on Γ_{ij}, we integrate by parts $\displaystyle\int_{\Gamma_{ij}} \phi_j^{\ell-1}\,\mathbf{q}_i\cdot\mathbf{n}_i$, introducing $\mathbf{q}_j^\star \in \mathbf{Q}_j$, a lifting of $\mathbf{q}_i\cdot\mathbf{n}_{i|\Gamma_{ij}}$ in $\mathbf{H}(\mathrm{div},\widetilde{\mathcal{R}}_j)$, such that $\mathbf{q}_j^\star = \nabla\,\psi_j^\star$ where $\psi_j^\star \in V_j^i$ (the space V_j^i is defined in (6.4.14)) solves the following elliptic problem:

$$\left\{\begin{array}{lcll}\text{Find } \psi_j^\star \in V_j^i \text{ such that} & & & \\ \Delta\psi_j^\star & = & c(\mathbf{q}_i) & \text{in } \widetilde{\mathcal{R}}_j \\ \partial_{\mathbf{n}_j}\psi_j^\star & = & -\mathbf{q}_i\cdot\mathbf{n}_{i|\Gamma_{ij}} & \text{on } \Gamma_{ij} \\ \partial_{\mathbf{n}_j}\psi_j^\star & = & 0 & \text{on } \partial\widetilde{\mathcal{R}}_j \setminus \overline{\Gamma_{ij}}\end{array}\right. .$$

Above, the constant $c(\mathbf{q}_i)$ is chosen so as to satisfy the compatibility condition of the Neumann problem, i.e. (cf. [97]), $\displaystyle\int_{\widetilde{\mathcal{R}}_j} c(\mathbf{q}_i) + \int_{\Gamma_{ij}} \mathbf{q}_i\cdot\mathbf{n}_i = 0$.
By construction, $\mathrm{div}\,\mathbf{q}_j^\star = c(\mathbf{q}_i)$ in $\widetilde{\mathcal{R}}_j$, and $\mathbf{q}_j^\star\cdot\mathbf{n}_j = -\mathbf{q}_i\cdot\mathbf{n}_i$ on Γ_{ij}, $\mathbf{q}_j^\star\cdot\mathbf{n}_j = 0$ on $\partial\widetilde{\mathcal{R}}_j \setminus \overline{\Gamma_{ij}}$. In the following, we propose a "natural" choice of discrete lifting. The integration by parts formula (4.2.1) in $\widetilde{\mathcal{R}}_j$ allows to write:

$$-\int_{\Gamma_{ij}} \phi_j^{\ell-1}\,\mathbf{q}_i\cdot\mathbf{n}_i = \int_{\Gamma_{ij}} \phi_j^{\ell-1}\,\mathbf{q}_j^\star\cdot\mathbf{n}_j = \int_{\partial\widetilde{\mathcal{R}}_j} \phi_j^{\ell-1}\,\mathbf{q}_j^\star\cdot\mathbf{n}_j = \int_{\widetilde{\mathcal{R}}_j} \nabla\,\phi_j^{\ell-1}\cdot\mathbf{q}_j^\star + \int_{\widetilde{\mathcal{R}}_j} \phi_j^{\ell-1}\,\mathrm{div}\,\mathbf{q}_j^\star,$$

$$= \int_{\widetilde{\mathcal{R}}_j} \left(-(D_j)^{-1}\mathbf{p}_j^{\ell-1}\cdot\mathbf{q}_j^\star + \phi_j^{\ell-1}\,\mathrm{div}\,\mathbf{q}_j^\star\right), \quad \text{from (6.5.5)-(i).}$$

Equation (6.5.8) now reads:

$$\begin{aligned}&\int_{\widetilde{\mathcal{R}}_i} \left((D_i)^{-1}\mathbf{p}_i^\ell \cdot \mathbf{q}_i - \phi_i^\ell \,\mathrm{div}\,\mathbf{q}_i\right) + \sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} (\alpha_{ij}^i)^{-1}\mathbf{p}_i^\ell\cdot\mathbf{n}_i\,\mathbf{q}_i\cdot\mathbf{n}_i\\ &= \sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} (\alpha_{ij}^i)^{-1}\mathbf{p}_j^{\ell-1}\cdot\mathbf{n}_j\,\mathbf{q}_j^\star\cdot\mathbf{n}_j + \sum_{j\in\mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\widetilde{\mathcal{R}}_j} \left(-(D_j)^{-1}\mathbf{p}_j^{\ell-1}\cdot\mathbf{q}_j^\star + \phi_j^{\ell-1}\,\mathrm{div}\,\mathbf{q}_j^\star\right).\end{aligned} \tag{6.5.9}$$

To take account of (6.5.5)-(0), we simply use the equivalent variational equality:

$$\forall \psi_i \in L^2(\widetilde{\mathcal{R}}_i), \quad 0 = \int_{\widetilde{\mathcal{R}}_i} (\operatorname{div} \mathbf{p}_i^\ell + \Sigma_{a,i}\, \phi_i^\ell - S_{f,i})\psi_i. \tag{6.5.10}$$

For $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, by subtracting the variational equalities (6.5.9) and (6.5.10), we conclude that if $(\mathbf{p}_i^\ell, \phi_i^\ell)$ is a solution to (6.5.5), then $(\mathbf{p}_i^\ell, \phi_i^\ell)$ is a solution to the variational formulation below:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{p}_i^\ell, \phi_i^\ell) \in \mathbf{X}_i^+ \text{ such that for all } (\mathbf{q}_i, \psi_i) \in \mathbf{X}_i^+ \\ \\ \displaystyle\int_{\widetilde{\mathcal{R}}_i} \left(-(D_i)^{-1}\mathbf{p}_i^\ell \cdot \mathbf{q}_i + \phi_i^\ell \operatorname{div} \mathbf{q}_i + \psi_i \operatorname{div} \mathbf{p}_i^\ell + \Sigma_{a,i}\, \phi_i^\ell \psi_i\right) \\ \qquad \displaystyle - \sum_{j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} (\alpha_{ij}^i)^{-1} \mathbf{p}_i^\ell \cdot \mathbf{n}_i\, \mathbf{q}_i \cdot \mathbf{n}_i \\ \displaystyle = \int_{\widetilde{\mathcal{R}}_i} S_{f,i}\psi_i - \sum_{j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\widetilde{\mathcal{R}}_j} \left(-(D_j)^{-1}\mathbf{p}_j^{\ell-1} \cdot \mathbf{q}_j^\star + \phi_j^{\ell-1} \operatorname{div} \mathbf{q}_j^\star\right) \\ \qquad \displaystyle - \sum_{j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} (\alpha_{ij}^i)^{-1} \mathbf{p}_j^{\ell-1} \cdot \mathbf{n}_j\, \mathbf{q}_j^\star \cdot \mathbf{n}_j. \end{array} \right. \tag{6.5.11}$$

In the following, for $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, we will use the bilinear forms $a_2^i(\cdot,\cdot)$, continuous on $\mathbf{X}_i^+ \times \mathbf{X}_i^+$:

$$a_2^i \ : \ ((\mathbf{q},\phi),(\mathbf{r},\psi)) \mapsto \int_{\widetilde{\mathcal{R}}_i} \left(-(D_i)^{-1}\mathbf{q}\cdot\mathbf{r} + \phi \operatorname{div}\mathbf{r} + \psi \operatorname{div}\mathbf{q} + \Sigma_{a,i}\,\phi\,\psi\right). \tag{6.5.12}$$

The bilinear form $\widetilde{a}_2^i(\cdot,\cdot)$ related to variational formulation (6.5.11) writes:

$$\left\{ \begin{array}{lcl} \widetilde{a}_2^i : \mathbf{X}_i^+ \times \mathbf{X}_i^+ & \to & \mathbb{R} \\ ((\mathbf{q},\phi),(\mathbf{r},\psi)) & \mapsto & a_2^i((\mathbf{q},\phi),(\mathbf{r},\psi)) - \displaystyle\sum_{j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}} \int_{\Gamma_{ij}} (\alpha_{ij}^i)^{-1} \mathbf{q}\cdot\mathbf{n}_i\, \mathbf{r}\cdot\mathbf{n}_i \quad . \end{array} \right. \tag{6.5.13}$$

Suppose that $\alpha_{ij}^i > 0$. According to Theorem 6.2.3, the bilinear form $\widetilde{a}_2^i(\cdot,\cdot)$ is T-coercive on $\mathbf{X}_i^+$, and Problem (6.5.11) is well posed.

6.5.2 Discretization

We define a family of triangulation $(\widetilde{\mathcal{R}}_{i,h})_h$ of $\widetilde{\mathcal{R}}_i$ made of N_i meshes such that: $\overline{\widetilde{\mathcal{R}}_{i,h}} = \cup_{l \in \mathcal{I}_i^T} R_{i,l}$ with $\mathcal{I}_i^T = \{1,\dots,N_i\}$; and F_i^+ faces $(f_{i,m})_{m \in \mathcal{I}_i^{f+}}$ with $\mathcal{I}_i^{f+} = \{1,\dots,F_i^+\}$. We set:

$$\mathcal{I}_i^f = \{m \in \mathcal{I}_i^{f+} \mid f_{i,m} \not\subset \Sigma_i\}, \quad \mathcal{I}_{i,ij}^f = \{m \in \mathcal{I}_i^{f+} \mid f_{i,m} \subset \Gamma_{ij}\}.$$

We suppose that for all $(i,j) \in \mathcal{I}\mathcal{J}$, triangulations $\widetilde{\mathcal{R}}_{i,h}$ and $\widetilde{\mathcal{R}}_{j,h}$ share their faces on Γ_{ij}, and we denote by $F_{ij} := |\mathcal{I}_{i,ij}^f| = |\mathcal{I}_{j,ij}^f|$.

We discretize $\mathbf{X}_i^+$ with $\mathrm{RTN}_{[k]}$ finite elements $\mathbf{X}_{i,h}^k := \mathbf{Q}_{i,h}^k \times L_{i,h}^k$, where $\mathbf{Q}_{i,h}^k$ and $L_{i,h}^k$ correspond to the spaces $\mathbf{Q}_h^k$ and L_h^k defined in §6.2.2 with $\mathcal{R}_h = \widetilde{\mathcal{R}}_{i,h}$, and $k \geq 0$. Let us set $(\underline{\psi}_{i,l})_{l \in \mathcal{I}_i}$ the basis functions of $L_{i,h}^k$, such that: $\underline{\psi}_{i,l|R_{i,l'}} = \delta_{l,l'}$. We denote by $(\underline{\omega}_{i,m})_{m \in \mathcal{I}_i^{f+}}$ the basis functions of $\mathbf{Q}_{i,h}^k$. For all $(i,j) \in \mathcal{I}\mathcal{J}$, triangulations $\widetilde{\mathcal{R}}_{i,h}$ and $\widetilde{\mathcal{R}}_{j,h}$ being identical on interface Γ_{ij},the spaces of the traces which are normal to Γ_{ij},

$$\mathfrak{T}_{i,ij}^f := \operatorname{span}\left((\underline{\omega}_{i,m}\cdot\mathbf{n}_{i|\Gamma_{ij}})_{m \in \mathcal{I}_{i,ij}^f}\right), \quad \mathfrak{T}_{j,ij}^f := \operatorname{span}\left((\underline{\omega}_{j,m}\cdot\mathbf{n}_{j|\Gamma_{ij}})_{m \in \mathcal{I}_{j,ij}^f}\right) \tag{6.5.14}$$

are the same: for all $m_i \in \mathcal{I}^f_{i,ij}$ it exists a unique $m_j \in \mathcal{I}^f_{j,ij}$ such that $\underline{\omega}_{i,m_i} \cdot \mathbf{n}_{i|\Gamma_{ij}} = -\underline{\omega}_{j,m_j} \cdot \mathbf{n}_{j|\Gamma_{ij}}$. The discrete liftings can then be chosen as follows:

$$\begin{array}{l} \forall (i,j) \in \mathcal{IJ},\ \forall (m_i,m_j) \in \mathcal{I}^f_{i,ij} \times \mathcal{I}^f_{j,ij} \text{ such that } f_{i,m_i} = f_{j,m_j} : \\ \text{- the basis function } \underline{\omega}_{j,m_j} \text{ is a lifting of } \underline{\omega}_{i,m_i} \cdot \mathbf{n}_{i|\Gamma_{ij}} \text{ in } \widetilde{\mathcal{R}}_j\,; \\ \text{- the basis function } \underline{\omega}_{i,m_i} \text{ is a lifting of } \underline{\omega}_{j,m_j} \cdot \mathbf{n}_{j|\Gamma_{ij}} \text{ in } \widetilde{\mathcal{R}}_i. \end{array} \tag{6.5.15}$$

Let us set $\mathbf{X}_{i,h} := \mathbf{X}^1_{i,h}$. We look for $(\mathbf{p}^\ell_{i,h}, \phi^\ell_{i,h})_{i\in\tilde{\mathcal{I}}_\mathcal{R}} \in \prod_{i\in\tilde{\mathcal{I}}_\mathcal{R}} \mathbf{X}_{i,h}$, an approximation of $(\mathbf{p}^\ell_i, \phi^\ell_i)_{i\in\tilde{\mathcal{I}}_\mathcal{R}}$ such that for all $i \in \tilde{\mathcal{I}}_\mathcal{R}$: $\mathbf{p}^\ell_{i,h} = \sum_{m\in\mathcal{I}^{a+}_i} P^\ell_{i,m}\,\underline{\omega}_{i,m}$ and $\phi^\ell_{i,h} = \sum_{l\in\mathcal{I}^T_i} \Phi^\ell_{i,l}\,w_{i,l}$.
The discretization of the variational formulation (6.5.11) writes:
Find $(\mathbf{p}^\ell_{i,h}, \phi^\ell_{i,h})_{i\in\tilde{\mathcal{I}}_\mathcal{R}} \in \prod_{i\in\tilde{\mathcal{I}}_\mathcal{R}} \mathbf{X}_{i,h}$ such that for all $i \in \tilde{\mathcal{I}}_\mathcal{R}$, for all $(l,m) \in \mathcal{I}^T_i \times \mathcal{I}^f_i$:

$$a^i_2((\mathbf{p}^\ell_{i,h}, \phi^\ell_{i,h})\,,\,(\underline{\omega}_{i,m}, \underline{\psi}_{i,l})) = \int_{\widetilde{\mathcal{R}}_i} S_{f,i}\,\underline{\psi}_{i,l}, \tag{6.5.16}$$

and for all $i \in \tilde{\mathcal{I}}_\mathcal{R}$, for all $j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}$, for all $(l,m_i) \in \mathcal{I}^T_i \times \mathcal{I}^f_{i,ij}$:

$$\begin{aligned} &a^i_2((\mathbf{p}^\ell_{i,h}, \phi^\ell_{i,h})\,,\,(\underline{\omega}_{i,m_i}, \underline{\psi}_{i,l})) - \int_{\Gamma_{ij}} (\alpha^i_{ij})^{-1} \mathbf{p}^\ell_{i,h}\cdot\mathbf{n}_i\,\underline{\omega}_{i,m_i}\cdot\mathbf{n}_i \\ &= \int_{\widetilde{\mathcal{R}}_i} S_{f,i}\,\underline{\psi}_{i,l} - \int_{\widetilde{\mathcal{R}}_j} \left(-(D_j)^{-1}\mathbf{p}^{\ell-1}_{j,h}\cdot\underline{\omega}_{j,m_j} + \phi^{\ell-1}_{j,h}\ \mathrm{div}\,\underline{\omega}_{j,m_j}\right) \\ &\qquad - \int_{\Gamma_{ij}} (\alpha^i_{ij})^{-1}\mathbf{p}^{\ell-1}_{j,h}\cdot\mathbf{n}_j\,\underline{\omega}_{j,m_j}\cdot\mathbf{n}_j. \end{aligned} \tag{6.5.17}$$

This algorithm corresponds to a block Jacobi to solve (A.3.1). In Appendix A.3, we express algorithm (6.5.17) with matrices.

6.5.3 Numerical illustration

The results presented here concern a pressurized water reactor core of capacity 900 MWe ($3D$ PWR 900 MWe core). The cross sections that we use come from experimental results [182] and they are highly heterogeneous. We solve Problem (6.1.12) (SP_1 approximation) with two groups of energy (cf. $G = 1$ in §6.1.2). The domain decomposition algorithm substitute the space solver in the inverse power algorithm (A.1). The α-parameter is set using Equation (6.5.6). It may differ from one group of energy to another and from one cell to another, since the cross-sections are highly heterogeneous. In Table 6.1, we present the results for computations of the MINOS solver from 1 to 128 computational cores with RT_0 finite elements. The data of this table is:

- Nc: The number of cores (ie. the number of subdomains).
- N_{DD}: The $3D$ Cartesian (N^x_{DD}, N^y_{DD} ,N^z_{DD}) decomposition.
- N_{out}: The number of iterations of the inverse power algorithm to converge.
- $Err.$: The (unsigned) difference between the computed and the converged eigenvalues, either sequentially or in parallel, times 10^{-5}.
- CPU: The CPU time spent within the MINOS solver, given in seconds.
- $Eff.$: The efficiency (in %): namely, $T_1/(N_{DD} \times T_N)$, where T_1 (resp. T_N) is the total sequential (resp. parallel, on N_{DD} cores) CPU time.

The method scales relatively well: its efficiency ranges from 67% to 88%. The number of outer iterations does not increase much, and accuracy is steady.
More details are given in [A11, §7]. We also studied the use of both GPU and domain decomposition in [P8]. Additional results are presented in [A10, P6], where we give details on the optimized Schwarz method applied to the SP_N approximation (6.1.6).

Nc	N_{DD} (x,y,z)	N_{out}	$Err.\ 10^{-5}$	CPU (s)	$Eff.$ (%)
1	$(1,1,1)$	380	0	284	100
2	$(2,1,1)$	380	0	166	85
4	$(2,2,1)$	379	0	105	67
8	$(2,2,2)$	377	0	53	67
16	$(4,4,1)$	377	0	25	71
32	$(4,4,2)$	385	0	10	88
64	$(8,8,1)$	393	0	5	88
128	$(8,8,2)$	393	0	2.6	85

Table 6.1: Computations results with the proposed numerical scheme.

The Optimized Schwarz method is simple to implement and it has proven its effectiveness. Nevertheless, for some nuclear reactor core configurations, it would be usefull to manage local refinement (cf. Figure 6.2(b)). To this end, we have developed a domain decomposition method in which interface conditions are handled using Lagrange multipliers.

6.6 DD+L^2-jumps method [A6, A2, P5]

Again, we wish to solve Problem (6.4.13). The idea developed below is that, rather than considering the normal trace of $\mathbf{p}$ on each side of the interfaces separately, we are interested directly in the difference between these traces on the interfaces (what we have called the jump). We need a Lagrange multiplier ϕ_S representing the trace of ϕ on Γ_S belonging to $H^{\frac{1}{2}}_{-}(\Gamma_S)$

$$H^{\frac{1}{2}}_{-}(\Gamma_S) = \left\{ \psi_S \in M \,|\, \psi_{S,ij} \in H^{\frac{1}{2}}(\Gamma_{ij}),\, \forall (i,j) \in \mathcal{IJ} \right\}, \tag{6.6.1}$$

where we recall that $M = \left\{ \psi_S \in \prod_{(i,j)\in\mathcal{IJ}} L^2(\Gamma_{ij}) \right\}$ and where we set $\psi_{S,ij} := \psi_{S|\Sigma_{ij}}$. We endow $H^{\frac{1}{2}}_{-}(\Gamma_S)$ with the graph norm.
Problem (6.4.12) is equivalent to Problem (6.6.2) below:

$$\left\{ \begin{array}{lrcll} \text{Find } ((\mathbf{p},\phi),\phi_S) \in (\widetilde{\mathcal{P}}\mathbf{H}(\mathrm{div},\mathcal{R}) \times \widetilde{\mathcal{P}}V) \times H^{\frac{1}{2}}_{-}(\Gamma_S) \text{ such that:} & & & & \\ \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}} & \mathrm{div}\,\mathbf{p}_i + \Sigma_{a,i}\phi_i & = & S_{f,i} & \text{in } \widetilde{\mathcal{R}}_i \quad (0) \\ \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}} & (D_i)^{-1}\mathbf{p}_i + \nabla\,\phi_i & = & 0 & \text{in } \widetilde{\mathcal{R}}_i \quad (\mathrm{i}) \\ \forall i \in \tilde{\mathcal{I}}_{\mathcal{R}}\, \forall j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}, & \phi_i & = & \phi_S & \text{on } \Sigma_{ij} \quad (\mathrm{ii}) \\ \forall (i,j) \in \mathcal{IJ} & [\mathbf{p}\cdot\mathbf{n}]_{ij} & = & 0 & \text{on } \Sigma_{ij} \quad (\mathrm{iii}) \end{array} \right. . \tag{6.6.2}$$

We introduce finally the following Hilbert spaces (the space $\widetilde{\mathbf{Q}}$ is defined in (6.4.10)):

$$\widetilde{\mathbf{X}} = \left\{ \xi := (\mathbf{q},\psi) \in \widetilde{\mathbf{Q}} \times L^2(\mathcal{R}) \right\}, \|\xi\|_{\widetilde{\mathbf{X}}} := \left(\|\mathbf{q}\|^2_{\widetilde{\mathbf{Q}}} + \|\psi\|^2_{L^2(\mathcal{R})} \right)^{1/2};$$

$$\mathtt{W} = \left\{ \mathtt{w} := (\xi,\psi_S) \in \widetilde{\mathbf{X}} \times M \right\}, \|\mathtt{w}\|_{\mathtt{W}} := \left(\|\xi\|^2_{\widetilde{\mathbf{X}}} + \|\psi_S\|^2_M \right)^{1/2}.$$

6.6.1 Variational Formulation

On the one hand $\phi \in L^2(\mathcal{R})$. On the other hand, we notice, cf. (6.4.8), that $[\mathbf{p}\cdot\mathbf{n}]_{ij} \in H'_{ij}$ for all $(i,j) \in \mathcal{I} \times \mathcal{I}_{\widetilde{\mathcal{R}}_i}$. Morevover, $L^2(\Sigma_{ij}) \subset H'_{ij}$, cf. §4.1. Noticing that $0 \in L^2(\Sigma_{ij})$, we conclude from, (6.6.2)-(iii), that in particular, $[\mathbf{p}\cdot\mathbf{n}]_{ij} \in L^2(\Sigma_{ij})$ for all $(i,j) \in \mathcal{I} \times \mathcal{I}_{\widetilde{\mathcal{R}}_i}$, and $[\mathbf{p}\cdot\mathbf{n}] \in M$. Thus, $\mathbf{p} \in \widetilde{\mathbf{Q}}$, where $\widetilde{\mathbf{Q}}$ is defined in (6.4.10) and we have to find $(\mathbf{p},\phi) \in \widetilde{\mathbf{Q}} \times L^2(\mathcal{R})$. We then build a new variational formulation. Recall that, since Problem (6.4.12) is equivalent to Problem (6.4.1) or (6.4.2), $\phi \in H^1_0(\mathcal{R})$. We can use first the integration by parts formula (6.4.11), for $\mathbf{q} \in \widetilde{\mathbf{Q}}$, and then (6.6.2)-(ii) to identify the trace of ϕ on Γ_S, obtaining:

$$\int_{\mathcal{R}} \nabla\,\phi\cdot\mathbf{q} = -\int_{\mathcal{R}} \phi\ \mathrm{div}\,\mathbf{q} + \int_{\Gamma_S} [\mathbf{q}\cdot\mathbf{n}]\ \phi_S.$$

Using (6.6.2)-(i), we get:

$$\int_{\mathcal{R}} \nabla\,\phi\cdot\mathbf{q} = \sum_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} \int_{\widetilde{\mathcal{R}}_i} \nabla\,\phi_i\cdot\mathbf{q}_i = -\sum_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} \int_{\widetilde{\mathcal{R}}_i} (D_i)^{-1}\mathbf{p}_i\cdot\mathbf{q}_i = -\int_{\mathcal{R}} D^{-1}\mathbf{p}\cdot\mathbf{q}.$$

Thus, for all $\mathbf{q} \in \widetilde{\mathbf{Q}}$, it holds: $-\displaystyle\int_{\mathcal{R}} D^{-1}\mathbf{p}\cdot\mathbf{q} + \int_{\mathcal{R}} \phi\ \mathrm{div}\,\mathbf{q} - \int_{\Gamma_S} [\mathbf{q}\cdot\mathbf{n}]\ \phi_S = 0$. Moreover, for all $v \in L^2(\mathcal{R})$, using (6.6.2)-(0), we have: $\displaystyle\int_{\mathcal{R}} (\mathrm{div}\,\mathbf{p} + \Sigma_a\phi)\psi = \int_{\mathcal{R}} S_f\psi$. Finally, for all $\psi_S \in M$, from (6.6.2)-(iii) it follows that: $\displaystyle\int_{\Gamma_S} [\mathbf{p}\cdot\mathbf{n}]\ \psi_S = 0$. To conclude, if the triplet $(\mathbf{p},\phi,\phi_S)$ is solution to (6.6.2), then it solves the variational formulation:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{p},\phi,\phi_S) \in \widetilde{\mathbf{Q}} \times L^2(\mathcal{R}) \times M \text{ such that } \forall (\mathbf{q},\psi,\psi_S) \in \widetilde{\mathbf{Q}} \times L^2(\mathcal{R}) \times M : \\ \displaystyle\int_{\mathcal{R}} \left(-D^{-1}\mathbf{p}\cdot\mathbf{q} + \phi\ \mathrm{div}\,\mathbf{q} + \psi\,\mathrm{div}\,\mathbf{p} + \Sigma_a\phi\,\psi\right) - \int_{\Gamma_S} \left([\mathbf{q}\cdot\mathbf{n}]\ \phi_S + [\mathbf{p}\cdot\mathbf{n}]\ \psi_S\right) = \int_{\mathcal{R}} S_f\psi \end{array}\right. . \tag{6.6.3}$$

Above, ϕ_S, ψ_S play the role of Lagrange multipliers, with M the space of those Lagrange multipliers. We call the mixed, multi-domain variational formulation (6.6.3) the *domain decomposition+L^2-jumps method* (or DD+L^2-jumps method). From now on, we use the notations:

- $\mathtt{u} = (\zeta,\phi_S)$, $\zeta = (\mathbf{p},\phi)$, $\mathbf{p} = (\mathbf{p}_i)_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}$ and $\phi = (\phi_i)_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}$;
- $\mathtt{w} = (\xi,\psi_S)$, $\xi = (\mathbf{q},\psi)$, $\mathbf{q} = (\mathbf{q}_i)_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}$ and $\psi = (\psi_i)_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}$;

and we define the bilinear forms:

$$c : \mathtt{W}\times\mathtt{W} \to \mathbb{R}, \quad c(\mathtt{u},\mathtt{w}) = \int_{\mathcal{R}} \left(-D^{-1}\,\mathbf{p}\cdot\mathbf{q} + \phi\ \mathrm{div}\,\mathbf{q} + \psi\ \mathrm{div}\,\mathbf{p} + \Sigma_a\,\phi\,\psi\right), \tag{6.6.4}$$

$$\ell_S : \mathtt{W}\times\mathtt{W} \to \mathbb{R}, \qquad \ell_S(\mathtt{u},\mathtt{w}) = \int_{\Gamma_S} [\mathbf{p}\cdot\mathbf{n}]\,\psi_S, \tag{6.6.5}$$

$$c_S : \mathtt{W}\times\mathtt{W} \to \mathbb{R}, \qquad c_S(\mathtt{u},\mathtt{w}) = c(\mathtt{u},\mathtt{w}) - \ell_S(\mathtt{u},\mathtt{w}) - \ell_S(\mathtt{w},\mathtt{u}). \tag{6.6.6}$$

Let $f \in \mathcal{L}(\mathtt{W},\mathbb{R})$ such that for all $wt \in \mathtt{W}$ $f(\mathtt{w}) = \displaystyle\int_{\mathcal{R}} S_f\psi$. We write the DD+$L^2$-jumps method (6.6.3) as:

$$\text{Find } \mathtt{u} \in \mathtt{W} \text{ such that } \forall \mathtt{w} \in \mathtt{W}\text{:}\quad c_S(\mathtt{u},\mathtt{w}) = f(\mathtt{w}). \tag{6.6.7}$$

Notice that $c_S(\cdot,\cdot)$ is symmetric. We can prove that solution to (6.6.7) satisfies (6.6.2), cf. [A6, Theorem 2]. In order to prove the well-posedness of (6.6.7), we need to define a lifting of the Lagrange multipliers. Consider $\phi^\star_S \in M$.

- If $i \in \mathcal{I}_S$ ($\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S = \partial\widetilde{\mathcal{R}}_i$), we let $\widetilde{\phi}_i = -\dfrac{1}{2}|\widetilde{\mathcal{R}}_i|^{-1} \displaystyle\int_{\partial\widetilde{\mathcal{R}}_i} \phi^\star_S$ and $\widetilde{\phi}_{i,S} := \phi^\star_{S|\partial\widetilde{\mathcal{R}}_i}$.

- If $i \notin \mathcal{I}_S$, we let $\widetilde{\phi}_i = 0$. We consider $\widetilde{\phi}_{i,S} \in L^2(\partial\widetilde{\mathcal{R}}_i)$, the continuation of $\phi^\star_{S|\partial\widetilde{\mathcal{R}}_i\cap\Gamma_S}$ by a constant value $\gamma_{i,S}$ to $\partial\widetilde{\mathcal{R}}_i$. The constant $\gamma_{i,S}$ is chosen so that the mean value of $\widetilde{\phi}_{i,S}$ vanishes.

For $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, one has $\widetilde{\phi}_i \in L^2(\widetilde{\mathcal{R}}_i)$ and $\widetilde{\phi}_{i,S} \in L^2(\partial\widetilde{\mathcal{R}}_i)$ with: $\|\widetilde{\phi}_i\|_{L^2(\widetilde{\mathcal{R}}_i)} + \|\widetilde{\phi}_{i,S}\|_{L^2(\partial\widetilde{\mathcal{R}}_i)} \lesssim \|\phi^\star_S\|_{L^2(\partial\widetilde{\mathcal{R}}_i\cap\Gamma_S)}$. We choose to solve the equivalent problem:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{v}_i, u_i) \in \mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i) \times H^1(\widetilde{\mathcal{R}}_i) \text{ such that} \\ \begin{array}{rcll} -\mathbf{v}_i + \nabla u_i &=& 0, & \text{in } \widetilde{\mathcal{R}}_i \\ -\operatorname{div}\mathbf{v}_i &=& \widetilde{\phi}_i, & \text{in } \widetilde{\mathcal{R}}_i \\ \mathbf{v}_i \cdot \mathbf{n}_i &=& \dfrac{1}{2}\widetilde{\phi}_{i,S} & \text{on } \partial\widetilde{\mathcal{R}}_i \end{array} \end{array}\right. . \tag{6.6.8}$$

Well-posedness and continuity of the lifting follow, indeed for the latter we have

$$\|\mathbf{v}_i\|_{\mathbf{H}(\mathrm{div},\widetilde{\mathcal{R}}_i)} \leq C_{lift,i}\, \|\phi^\star_S\|_{L^2(\partial\widetilde{\mathcal{R}}_i\cap\Gamma_S)}. \tag{6.6.9}$$

Regarding now the regularity of $\mathbf{v}_i$, there holds $\mathbf{v}_i \in \mathbf{H}^{1/2}(\widetilde{\mathcal{R}}_i)$, and $\|\mathbf{v}_i\|_{\mathbf{H}^{1/2}(\widetilde{\mathcal{R}}_i)} \lesssim \|\phi^\star_S\|_{L^2(\partial\widetilde{\mathcal{R}}_i\cap\Gamma_S)}$. An equivalent, mixed variational formulation of Problem (6.6.8) reads, cf. [50, §7]:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{v}_i, u_i) \in \mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i) \times L^2(\widetilde{\mathcal{R}}_i) \text{ such that } \forall\, (\mathbf{q}^0_i, \psi_i) \in \mathbf{H}_0(\mathrm{div}, \widetilde{\mathcal{R}}_i) \times L^2(\widetilde{\mathcal{R}}_i): \\ \begin{array}{rcl} \displaystyle\int_{\widetilde{\mathcal{R}}_i} \mathbf{v}_i \cdot \mathbf{q}^0_i + \int_{\widetilde{\mathcal{R}}_i} \operatorname{div}\mathbf{v}_i\, \psi_i + \int_{\widetilde{\mathcal{R}}_i} \operatorname{div}\mathbf{q}^0_i\, u_i &=& \displaystyle -\int_{\widetilde{\mathcal{R}}_i} \widetilde{\phi}_i\, \psi_i, \\ \mathbf{v}_i \cdot \mathbf{n}_i &=& \dfrac{1}{2}\widetilde{\phi}_{i,S} \text{ on } \partial\widetilde{\mathcal{R}}_i. \end{array} \end{array}\right. \tag{6.6.10}$$

One actually solves the problem in the variable $\mathbf{v}^0_i := \mathbf{v}_i - \mathbf{v}^{lift}_i \in \mathbf{H}_0(\mathrm{div}, \widetilde{\mathcal{R}}_i)$, where $\mathbf{v}^{lift}_i \in \mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)$ is a continuous lifting of the normal trace $\dfrac{1}{2}\widetilde{\phi}_{i,S}$, and we can use basic T-coercivity to prove the well-posedness of Problem (6.6.7).

Theorem 6.6.1 (`[A6, Theorem 3]`). *Under assumptions given in* (6.1.7)*, the continuous bilinear form* $c_S(\cdot,\cdot)$ *is* T*-coercive on* $\mathtt{W} \times \mathtt{W}$*. The variational formulation* (6.6.7) *is then well posed. In particular, it exists a constant* $C > 0$ *depending only on* $\mathcal{R}$*,* Σ_a *and* D *such that for all* $S_f \in L^2(\mathcal{R})$*, it exists a unique tuple* $((\mathbf{p}, \phi), \phi_S)$ *solution to (6.6.7) such that*

$$\|\mathbf{p}\|_{\widetilde{\mathbf{Q}}} + \|\phi\|_{L^2(\mathcal{R})} + \|\phi_S\|_{L^2(\Gamma_S)} \leq C\, \|S_f\|_{L^2(\mathcal{R})}.$$

Proof. For all $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, we consider $(\mathbf{v}_i, u_i)$ the solution to Problem (6.6.10). Consider $T \in \mathcal{L}(\mathtt{W})$ such that $T((\mathbf{p}, \phi, \phi_S)) = (\mathbf{q}, \psi, \psi_S)$ where:

$$\left\{\begin{array}{rcll} \mathbf{q}_i &=& -\mathbf{p}_i - \alpha\mathbf{v}_i & \in \mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i), \quad \text{for } i \in \tilde{\mathcal{I}}_{\mathcal{R}}\,, \\ \psi_i &=& \dfrac{1}{2}\phi_i + \dfrac{1}{2}(\Sigma_{a,i})^{-1} \operatorname{div}\mathbf{p}_i & \in L^2(\widetilde{\mathcal{R}}_i), \quad \text{for } i \in \tilde{\mathcal{I}}_{\mathcal{R}}\,, \\ \psi_S &=& \phi_S - [\mathbf{p}\cdot\mathbf{n}] & \in M. \end{array}\right.$$

The coefficient α depends on $C_{lift,i}$, on the ratio $\max_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}(|\partial\widetilde{\mathcal{R}}_i|\,|\widetilde{\mathcal{R}}_i|)_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}$ and on the data. ◇

6.6.2 Discretization

The discretization of the product space $\widetilde{\mathbf{Q}} \times L^2(\mathcal{R})$ is the product space $\widetilde{\mathbf{Q}}_h \times M_h$ where $\widetilde{\mathbf{Q}}_h := \prod_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} \mathbf{Q}^{k_i}_{i,h}$ and $M_h := \prod_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}} L^{k_i}_{i,h}$, where $\mathbf{Q}^{k_i}_{i,h}$ and $L^{k_i}_{i,h}$ correspond here to the spaces $\mathbf{Q}^{k_i}_h$ and $L^{k_i}_h$ defined in §6.2.2 with $\mathcal{R}_h = \widetilde{\mathcal{R}}_{i,h}$, and $k_i \geq 0$[5]. Using a Lagrange multiplier, the faces of two neighbouring subdomains may not match, and the approximation order of two neighbouring subdomains may differ. Let $T_{i,h}$ be

[5] Compare to the optimized Schwarz decomposition method, we can choose different order approximation in each subdomain.

the space of the normal traces of elements of $\mathbf{Q}_{i,h}^{k_i}$ on $\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S$:

$$T_{i,h} := \left\{ q_{i,h} \in L^2(\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S) \,|\, \exists\, \mathbf{q}_{i,h} \in \mathbf{Q}_{i,h}^{k_i} \,|\, q_{i,h} = \mathbf{q}_{i,h} \cdot \mathbf{n}_{i|\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S} \right\}. \tag{6.6.11}$$

notice that $T_{i,h|\Gamma_{ij}} = \mathfrak{T}_{i,ij}^f$, defined by Equation (6.5.14). Several situations[6] can occur on a given interface Γ_{ij}:

(1) *non-nested meshes*: $T_{i,h} \not\subset T_{j,h}$ and $T_{j,h} \not\subset T_{i,h}$;

(2) *nested meshes*: $T_{i,h} \subset T_{j,h}$ or $T_{j,h} \subset T_{i,h}$;

(3) *matching meshes*: nested meshes with $T_{i,h} = T_{j,h}$.

Remark 6.6.1. *Matching domain decompositions correspond to matching meshes ; resp., non-matching domain decompositions correspond either to non-nested, or nested but non-matching, meshes.*

Usually, the term *nested meshes* is used to describe a family of successively refined meshes. In this paper, we will use this expression to express that on all interfaces Γ_{ij}, case (2) described above holds.

Let us denote by $M_h \subset M$ the discrete space of the Lagrange multipliers, and by $M_h^{ij} \subset L^2(\Gamma_{ij})$ the discrete space of the Lagrange multipliers restricted to the interface Γ_{ij}. We introduce the discrete projection operators from the spaces of normal traces $T_{i,h}$ to M_h, and vice versa:

$$\left\{ \begin{array}{rcl} \Pi_{S,i} : T_{i,h} & \to & M_h \\ q_{i,h} & \mapsto & \Pi_{S,i}(q_{i,h}) \end{array} \right. , \qquad \left\{ \begin{array}{rcl} \pi_{S,i} : M_h & \to & T_{i,h} \\ \psi_{S,h} & \mapsto & \pi_{S,i}(\psi_{S,h}) \end{array} \right. .$$

These projections are defined by:

$$\forall q_{i,h} \in T_{i,h},\, \forall \psi_{S,h} \in M_h \quad \left\{ \begin{array}{l} \displaystyle\int_{\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S} (\Pi_{S,i}(q_{i,h}) - q_{i,h})\ \psi_{S,h} = 0 \\ \displaystyle\int_{\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S} (\pi_{S,i}(\psi_{S,h}) - \psi_{S,h})\ q_{i,h} = 0 \end{array} \right. . \tag{6.6.12}$$

As the operators $\Pi_{S,i}$ and $\pi_{S,i}$ are orthogonal projections, they are continuous, with a continuity modulus equal to 1.

Proposition 6.6.1. *The following inequalities hold:*

$$\forall q_{i,h} \in T_{i,h},\, \forall \psi_{S,h} \in M_h \quad \left\{ \begin{array}{l} \|\pi_{S,i}(\psi_{S,h})\|_{L^2(\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S)} \leq \|\psi_{S,h}\|_{L^2(\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S)} \\ \|\Pi_{S,i}(q_{i,h})\|_{L^2(\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S)} \leq \|q_{i,h}\|_{L^2(\partial\widetilde{\mathcal{R}}_i \cap \Gamma_S)} \end{array} \right. . \tag{6.6.13}$$

We also introduce the orthogonal projection operator $\Pi_S^0 : M \to M_h^0$, where M_h^0 is the subspace of piecewise constant fields. According to [146, Proposition 1.135], letting h_S be the meshsize on Γ_S, we have:

$$\forall \psi_S \in H_-^{1/2}(\Gamma_S),\ \|\psi_S - \Pi_S^0(\psi_S)\|_M \lesssim h_S^{1/2} \|\psi_S\|_{H_-^{1/2}(\Gamma_S)}. \tag{6.6.14}$$

Next, let $\mathbf{p}_h \in \widetilde{\mathbf{Q}}_h$. We define the discrete jump of the normal component of $\mathbf{p}_h$ on the interface Γ_{ij} as $[\mathbf{p}_h \cdot \mathbf{n}]_{h,ij} := \sum_{l=i,j} \Pi_{S,l}(\mathbf{p}_{l,h} \cdot \mathbf{n}_{l|\Gamma_{ij}})$. The discrete global jump of the normal component, $[\mathbf{p}_h \cdot \mathbf{n}]_h \in M_h$, is defined by: $[\mathbf{p}_h \cdot \mathbf{n}]_{h|\Gamma_{ij}} := [\mathbf{p}_h \cdot \mathbf{n}]_{h,ij}$, for $i, j \in \mathcal{I}$. We finally define:

$$\widetilde{\mathbf{X}}_h = \left\{ \xi_h := (\mathbf{q}_h, \psi_h) \in \widetilde{\mathbf{Q}}_h \times L_h \right\}, \text{ endowed with } \|\cdot\|_{\widetilde{\mathbf{X}}},$$

$$\mathtt{W}_h = \left\{ \mathtt{w}_h := (\xi_h, \psi_{S,h}) \in \widetilde{\mathbf{X}}_h \times M_h \right\}, \text{ endowed with } \|\cdot\|_{\mathtt{W}}.$$

[6] Non-nested meshes are also called sliding meshes in the literature, see [246].

In our setting, the conforming discretization of the variational formulation (6.6.7) reads:
Find $((\mathbf{p}_h,\phi_h),\phi_{S,h})\in \mathtt{W}_h$, such that $\forall((\mathbf{q}_h,\psi_h),\psi_{S,h})\in \mathtt{W}_h$:

$$\begin{aligned}\int_{\mathcal{R}} \big(-D^{-1}\,\mathbf{p}_h\cdot\mathbf{q}_h+\phi_h\,\operatorname{div}\mathbf{q}_h+\psi_h\,\operatorname{div}\mathbf{p}_h+\sigma\,\phi_h\,\psi_h\big)\\ -\int_{\Gamma_S}[\mathbf{p}_h\cdot\mathbf{n}]\,\psi_{S,h}-\int_{\Gamma_S}[\mathbf{q}_h\cdot\mathbf{n}]\,\phi_{S,h}=\int_{\mathcal{R}} S_f\,\psi_h.\end{aligned}\tag{6.6.15}$$

Or equivalently:

$$\text{Find } \mathtt{u}_h\in\mathtt{W}_h \text{ such that } \forall \mathtt{w}_h\in\mathtt{W}_h,\ c_S(\mathtt{u}_h,\mathtt{w}_h)=f(\mathtt{w}_h).\tag{6.6.16}$$

For the study of the well-posedness in the discrete case, we have to lift continuously the discrete Lagrange multipliers. To that aim, one simply discretizes the variational formulation (6.6.10). So we assume that, given $\phi_{S,h}\in M_h$, there exists for all $i\in\tilde{\mathcal{I}}_{\mathcal{R}}$ a discrete lifting $\mathbf{v}_{i,h}\in\mathbf{Q}_{i,h}^{k_i}$ that satisfies the discrete counterpart of (6.6.9) and of (6.6.10) (the constant is denoted by C'_{lift}). We label those two discrete conditions (6.6.9h)-(6.6.10h). We can use discrete T-coercivity to prove the well-posedness of Problem (6.6.16).

Theorem 6.6.2 (`[A6, Theorem 4]`). *Assume that the conditions*

$$\exists\beta_h>0,\ \forall\mathbf{q}_h\in\widetilde{\mathbf{Q}}_h,\ \int_{\Gamma_S}[\mathbf{q}_h\cdot\mathbf{n}]_h\,[\mathbf{q}_h\cdot\mathbf{n}]\geq\beta_h\int_{\Gamma_S}[\mathbf{q}_h\cdot\mathbf{n}]^2\tag{6.6.17}$$

and

$$\exists\gamma_h>0,\ \forall\psi_{S,h}\in M_h,\quad \sum_{(i,j)\in\mathcal{I}\mathcal{J}}\int_{\Gamma_{ij}}\big(\pi_{S,i}(\psi_{S,h})^2+\pi_{S,j}(\psi_{S,h})^2\big)\geq\gamma_h\|\psi_{S,h}\|_M^2\tag{6.6.18}$$

hold, then the bilinear form $c_S(\cdot,\cdot)$ is T-coercive on $\mathtt{W}_h\times\mathtt{W}_h$. If in addition β_h and γ_h can be chosen independently of h in these two inequalities, the constant is independent of h.

Remark 6.6.2. *In other words, (6.6.17)-(6.6.18) can be seen as the* algebraic conditions. *We discuss these conditions in the next subsection.*

Below, we also use the orthogonal projection operators $\pi_i^0: L^2(\widetilde{\mathcal{R}}_i)\to L^0_{i,h}$ (see (6.2.6)).

Proof. Consider $\mathtt{u}_h:=((\mathbf{p}_h,\phi_h),\phi_{S,h})\in\mathtt{W}_h$, and define $T_h(\mathtt{u}_h)=\mathtt{w}_h:=((\mathbf{q}_h,\psi_h),\psi_{S,h})\in\mathtt{W}_h$ as

$$\left\{\begin{array}{lcll}\mathbf{q}_{i,h} &=& -\mathbf{p}_{i,h}-\alpha'\mathbf{v}_{i,h}, & \text{for } i\in\tilde{\mathcal{I}}_{\mathcal{R}},\\ \psi_{i,h} &=& \dfrac{1}{2}\phi_{i,h}+\dfrac{1}{2}\pi_i^0((\Sigma_{a,i})^{-1})\operatorname{div}\mathbf{p}_{i,h}, & \text{for } i\in\tilde{\mathcal{I}}_{\mathcal{R}},\\ \psi_{S,h} &=& \phi_{S,h}+[\mathbf{p}_h\cdot\mathbf{n}]_h, & \end{array}\right.\tag{6.6.19}$$

where α' depends on the ratio $\max_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}(|\partial\widetilde{\mathcal{R}}_i|\,|\widetilde{\mathcal{R}}_i^{-1})_{i\in\tilde{\mathcal{I}}_{\mathcal{R}}}$, on the data, on the constant C'_{lift} and on the constant γ_h. The discrete lifting $\mathbf{v}_{i,h}$ of $\phi_{S,h}$ is governed by (6.6.9h)-(6.6.10h) with data $\phi_S^\star=\pi_{S,i}(\phi_{S,h})$. ⋄

A straightforward consequence of (6.6.17) is obtained by taking $((\mathbf{q}_h,\psi_h),\psi_{S,h})=((0,0),-[\mathbf{p}_h\cdot\mathbf{n}]_h)\in\mathtt{W}_h$ as a test-function in (6.6.15), which leads to

$$0=\int_{\Gamma_S}[\mathbf{p}_h\cdot\mathbf{n}]\,[\mathbf{p}_h\cdot\mathbf{n}]_h\geq\beta_h\int_{\Gamma_S}[\mathbf{p}_h\cdot\mathbf{n}]^2.$$

Hence, $[\mathbf{p}_h\cdot\mathbf{n}]=0$ and so $\mathbf{p}_h\in\widetilde{\mathbf{Q}}_h\cap\mathbf{H}(\operatorname{div},\mathcal{R})$.

One infers the classical convergence result below [148, Lemma 27.17].

Corollary 6.6.1. *Assume that the algebraic conditions (6.6.17)-(6.6.18) are fulfilled uniformly in h, that the discretization $\widetilde{\mathbf{Q}}_h\times L_h$ satisfies approximability properties and that it is globally conforming in $\widetilde{\mathbf{Q}}\times L^2(\mathcal{R})$. Then the solutions $(\mathtt{u}_h)_h$ to the* DD+L^2-jumps method *converge to the solution $\mathtt{u}$ to (6.6.7). According to the first Strang's Lemma, and because $c_S(\cdot,\cdot)$ satisfies a uniform discrete inf-sup condition, the error reads:*

$$\|\mathtt{u}-\mathtt{u}_h\|_{\mathtt{W}}\lesssim\inf_{\mathtt{w}\in\mathtt{W}_h}\|\mathtt{u}-\mathtt{w}_h\|_{\mathtt{W}}.\tag{6.6.20}$$

The matrix system derived from (6.6.15) to be solved is detailed in [163, §7.1]. We propose next to study the conditions (6.6.17)-(6.6.18).

6.6.3 A study of the conditions (6.6.17)-(6.6.18)

Let us begin by some observations regarding the two conditions.
First, we noted that if (6.6.17) holds, then $[\mathbf{p}_h \cdot \mathbf{n}] = 0$ and $\mathbf{p}_h \in \widetilde{\mathbf{Q}}_h \cap \mathbf{H}(\mathrm{div}, \mathcal{R})$.
Second, if (6.6.18) holds, then $\psi_{S,h} \mapsto \left(\sum_{i,j\in\mathcal{IJ}} \int_{\Gamma_{ij}} (\pi_{S,i}(\psi_{S,h})^2 + \pi_{S,j}(\psi_{S,h})^2\right)^{1/2}$ is a norm on M_h. Indeed, we conclude from (6.6.13) that $\forall \psi_{S,h} \in M_h$

$$\gamma_h \|\psi_{S,h}\|_M^2 \leq \sum_{i,j\in\mathcal{IJ}} \int_{\Gamma_{ij}} \left(\pi_{S,i}(\psi_{S,h})^2 + \pi_{S,j}(\psi_{S,h})^2\right) \leq 2\|\psi_{S,h}\|_M^2.$$

To realize the conditions (6.6.17)-(6.6.18), one can think of two alternatives depending on whether the discrete space of the Lagrange multipliers M_h is build/inferred directly with the help of the discrete spaces of normal traces $(T_{i,h})_{i\in\tilde{\mathcal{I}}_\mathcal{R}}$. That is, whether $\sum_{i\in\tilde{\mathcal{I}}_\mathcal{R}} T_{i,h} \subset M_h$ or not.

The case $\sum_{i\in\tilde{\mathcal{I}}_\mathcal{R}} T_{i,h} \subset M_h$

One can disregard the situation where $\sum_{i\in\tilde{\mathcal{I}}_\mathcal{R}} T_{i,h}$ is a *strict* subset of M_h. In this case, there exists $\psi_{S,h}^{\perp} \neq 0$ that belongs to the orthogonal subspace of $\sum_{i\in\tilde{\mathcal{I}}_\mathcal{R}} T_{i,h}$ in M_h, and if $((\mathbf{p}_h, \phi_h), \phi_{S,h})$ is a solution to (6.6.15), then so is $((\mathbf{p}_h, \phi_h), \phi_{S,h} + \psi_{S,h}^{\perp})$. This is a straightforward consequence of the fact for all test-functions $((\mathbf{q}_h, \psi_h), \psi_{S,h})$, it holds $[\mathbf{q}_h \cdot \mathbf{n}] \in \sum_{i\in\tilde{\mathcal{I}}_\mathcal{R}} T_{i,h}$ and so $\int_{\Gamma_S} [\mathbf{q}_h \cdot \mathbf{n}]\, \psi_{S,h}^{\perp} = 0$: the discrete solution is not unique. Hence in order to obtain a well-posed problem, M_h must be chosen as:

$$M_h = \sum_{i\in\tilde{\mathcal{I}}_\mathcal{R}} T_{i,h}. \tag{6.6.21}$$

In this setting, conditions (6.6.17)-(6.6.18) follow readily with constants β_h and γ_h independent of h. In particular, all projections Π_i are equal to the identity so that :

$$\forall \mathbf{q}_h \in \widetilde{\mathbf{Q}}_h,\ \int_{\Gamma_S} [\mathbf{q}_h \cdot \mathbf{n}]_h\, [\mathbf{q}_h \cdot \mathbf{n}] = \int_{\Gamma_S} [\mathbf{q}_h \cdot \mathbf{n}]^2.$$

When M_h is defined by (6.6.21), there are two alternatives. For nested meshes on Γ_{ij}, the definition is straightforward, M_h^{ij} is the larger of the two spaces $T_{i,h}$ and $T_{j,h}$. On the other hand for non-nested meshes, the explicit construction of M_h^{ij} can be delicate. We refer to Gander et al [155, 156] for an implicit approach that allows to build the space M_h and the action of its test-fields in the variational formulation via an algorithm that is only used when the computations are performed.

Other constructions

To our knowledge, the other alternative has been first studied by Yotov [247] (in case of RTN finite elements). This is a situation where *nested* meshes are considered, but where $T_{i,h} + T_{j,h}$ and M_h^{ij} can be different on all interfaces Γ_{ij}. Here, the space M_h^{ij} must be taylored with care to obtain conditions (6.6.17)-(6.6.18). As an example, in [247], a $2D$ domain $\mathcal{R}$ was considered, with P_0 discrete normal traces with given meshsize $h_S = h_i$ on Γ_{ij}, P^1 discrete Lagrange multipliers with meshsize $2h_S$ and where the triangulation used for the normal traces is a subdivision of the triangulation used for the Lagrange multipliers. The best constant in (6.6.17) can be computed on the reference element, with a value of the parameter $\hat{\beta} = \frac{3}{4}$, so that $\beta_h = \frac{3}{4}$. Other examples can be constructed for $3D$ geometries. For instance, with Q_0 normal traces with given meshsize $h_S = h_i$ on Γ_{ij}, Q_1 discrete Lagrange multipliers

with meshsize $2h_S$ and the same rule governing the triangulations as above. In this case one finds that $\beta_h = \hat{\beta} = \frac{17}{32}$. Also, one readily checks that $\hat{\pi}\hat{\psi} = 0$ implies $\hat{\psi} = 0$ on the reference element, so condition (6.6.18) is fulfilled too with a constant γ_h that is independent of h. We postulate that these results could be extended as follows. Let $m \geq 2$: in $2D$ (resp. $3D$), P_0 (resp. Q_0) discrete normal traces with meshsize $h_S = h_i$, in conjunction with P_m (resp. Q_m) discrete Lagrange multipliers with meshsize $m\,h_S$ and again the same rule governing the triangulations as above.

6.6.4 A priori estimates

We consider here the case where for all $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$, $k_i = 0$. Let $\mathbf{q} \in \mathbf{H}(\mathrm{div}, \mathcal{R}) \cap \widetilde{\mathcal{P}}\mathbf{H}^{\mu}(\mathcal{R})$, with $0 < \mu$. A global RTN interpolant of $\mathbf{q}$ is defined on every subdomain $\widetilde{\mathcal{R}}_i$ via its restriction $\mathbf{q}_i$, and denoted by $\widetilde{\mathbf{q}}_{i,R}$ for $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$. One may thus define the global interpolant of $\mathbf{q}$ in $\widetilde{\mathbf{Q}}_h$, denoted by $\widetilde{\mathbf{q}}_R$ henceforth: $\widetilde{\mathbf{q}}_{R|\widetilde{\mathcal{R}}_i} = \widetilde{\mathbf{q}}_{i,R}$ for $i \in \tilde{\mathcal{I}}_{\mathcal{R}}$. Below, we also use the orthogonal projection operators $\pi_i^0 : L^2(\widetilde{\mathcal{R}}_i) \to L^0_{i,h}$ (see (6.2.6)) and $\Pi_S^0 : M \to M_h^0$ (see (6.6.14)). One has the following result:

Lemma 6.6.1 (`[A2, Lemma 6.1]`). *Assume that the meshes are nested, non-matching, on the interface Γ_{fc}, and that they are quasi-uniform on Γ_{fc}. To fix ideas, we assume $T_{c,h|\Gamma_{fc}} \subset T_{f,h|\Gamma_{fc}}$ with $T_{c,h|\Gamma_{fc}} \neq T_{f,h|\Gamma_{fc}}$ ([7]). Let $\mathbf{q} \in \mathbf{H}(\mathrm{div}, \mathcal{R}) \cap \mathbf{H}^{\mu}(\mathcal{R})$ with $0 < \mu < 1/2$, it holds:*

$$\|[\widetilde{\mathbf{q}}_R \cdot \mathbf{n}]\|_{L^2(\Gamma_{fc})} \lesssim h_f^{1/2} \,\|\, \mathbf{q}_f \,\|_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_f)}.$$

Theorem 6.6.3 (`[A2, Theorem 6.2]`). *Let the assumptions of Proposition 4.4.1 hold, with $r_{\max} < 1/2$. One has for matching meshes:*

$$\begin{array}{l} \forall \mu \in]0, r_{\max}[,\ \forall S_f \in H^{\mu}(\mathcal{R}), \\ \|\mathbf{p} - \mathbf{p}_h\|_{\mathbf{H}(\mathrm{div}, \mathcal{R})} + \|\phi - \phi_h\|_{L^2(\mathcal{R})} + \|\phi_S - \phi_{S,h}\|_M \lesssim h^{\mu}\, \|S_f\|_{H^{\mu}(\mathcal{R})}. \end{array} \tag{6.6.22}$$

For nested, non-matching *meshes, the result holds under the assumption that on an interface Γ_{ij} where the meshes $T_{i,h|\Gamma_{ij}}$ and $T_{j,h|\Gamma_{ij}}$ are non-matching ($T_{i,h|\Gamma_{ij}} \neq T_{j,h|\Gamma_{ij}}$), the families of triangulations of $T_{i,h|\Gamma_{ij}}$ and $T_{j,h|\Gamma_{ij}}$ are quasi-uniform.*

Proof. We bound the different contributions in the RHS of (6.6.20) for some appropriately chosen discrete field $\mathtt{w}_h$.
Matching meshes. We know that $[\mathbf{p} \cdot \mathbf{n}] = 0$. For matching meshes, one has also $[\widetilde{\mathbf{p}}_R \cdot \mathbf{n}] = 0$, so $[(\mathbf{p} - \widetilde{\mathbf{p}}_R) \cdot \mathbf{n}] = 0$. Starting from (6.6.20), the conclusion follows. Indeed, according to the a priori estimates (6.2.7), (6.2.10) and (6.6.14), $\mathtt{w}_h = (\widetilde{\mathbf{p}}_R, \pi^0\phi, \Pi_S^0(\phi_S)) \in \mathtt{W}_h$ is such that

$$\begin{aligned} \|\mathtt{u} - \mathtt{w}_h\|_{\mathtt{W}}^2 &= \sum_{i \in \tilde{\mathcal{I}}_{\mathcal{R}}} \|\mathbf{p}_i - \mathbf{p}_{i,R}\|^2_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)} + \|\phi - \pi^0\phi\|^2_{L^2(\mathcal{R})} + \|\phi_S - \Pi_S^0(\phi_S)\|^2_M \\ &\lesssim h^{2\mu}(|\mathbf{p}|^2_{\mathbf{H}^{\mu}(\mathcal{R})} + \|\,\mathrm{div}\,\mathbf{p}\|^2_{H^{\mu}(\mathcal{R})}) + h^2\|\phi\|^2_{\widetilde{\mathcal{P}}H^1(\mathcal{R})} + h_S\|\phi_S\|^2_{H_{-}^{1/2}(\Gamma_S)} \lesssim h^{2\mu}\,\|S_f\|^2_{H^{\mu}(\mathcal{R})}. \end{aligned}$$

Hence we conclude that for matching meshes it holds (recall that $\mathtt{u} = ((\mathbf{p}, \phi), \phi_S)$):

$$\|\mathtt{u} - \mathtt{u}_h\|_{\mathtt{W}} \lesssim h^{\mu}\,\|S_f\|_{H^{\mu}(\mathcal{R})}. \tag{6.6.23}$$

Nested meshes. In this case, $[\widetilde{\mathbf{p}}_R \cdot \mathbf{n}] \neq 0$ in general. Nonetheless, one can use the result of lemma 6.6.1, to find that

$$\|\,[(\mathbf{p} - \widetilde{\mathbf{p}}_R) \cdot \mathbf{n}]\,\|_M \lesssim h^{1/2}\,\|\mathbf{p}\|_{\mathbf{H}(\mathrm{div}, \mathcal{R})},$$

provided that the meshes are *quasi-uniform* on the part of the interface where they are non-matching. The estimate (6.6.23) still holds for nested meshes under this condition.
Conclusion. Noting that it always holds $[\mathbf{p} \cdot \mathbf{n}] = [\mathbf{p}_h \cdot \mathbf{n}] = 0$ (recall that $\mathbf{p}_h \in \widetilde{\mathbf{Q}}_h \cap \mathbf{H}(\mathrm{div}, \mathcal{R})$), developing the norm $\|\mathtt{u} - \mathtt{u}_h\|_{\mathtt{W}}$, one concludes:

$$\|\mathbf{p} - \mathbf{p}_h\|_{\mathbf{H}(\mathrm{div}, \mathcal{R})} + \|\phi - \phi_h\|_{L^2(\mathcal{R})} + \|\phi_S - \phi_{S,h}\|_M \lesssim h^{\mu}\,\|S_f\|_{H^{\mu}(\mathcal{R})}.$$

In other words, we have the a priori error estimate (6.6.22). ◇

As in the plain case, for "smooth data" S_f, i.e. $S_f \in H^{r_{\max}}(\mathcal{R})$, one expects a convergence rate at least in $h^{r_{\max}}$.

[7] f refers to *fine discretization*, while c refers to *coarse discretization*.

Remark 6.6.3. *Within our framework, we obtain error estimates that generalize those of [67, 245] for low-regularity solutions. In addition, the technical aspects we propose remain quite simple and natural since the discretization remains conforming (i.e.* $\widetilde{\mathbf{Q}}_h \times L_h \subset \widetilde{\mathbf{Q}} \times L^2(\mathcal{R})$*).*

To derive improved estimates on the error $\|\phi - \phi_h\|_{L^2(\mathcal{R})}$, we adapt the calculations of estimate (6.2.15) to the DD case. Let $\mathbf{Q}_h = \widetilde{\mathbf{Q}}_h \cap \mathbf{H}(\mathrm{div}, \mathcal{R})$ and $\mathbf{X}_h = \mathbf{Q}_h \times L_h$. We already know that when conditions (6.6.17)-(6.6.18) hold, the solution $((\mathbf{p}_h, \phi_h), \phi_{S,h}) \in \widetilde{\mathbf{X}}_h \times M_h$ of (6.6.16) (discrete DD case) is such that $(\mathbf{p}_h, \phi_h) \in \mathbf{X}_h$, since $\mathbf{p}_h \in \mathbf{Q}_h$. Then restricting the test-fields in (6.6.16) to elements of $\mathbf{X}_h \times M_h$ we observe that $(\mathbf{p}_h, \phi_h)$ satisfies (6.2.11) too (discrete plain-case), because all interface terms *vanish*. Hence, to estimate $\|\phi - \phi_h\|_{L^2(\mathcal{R})}$ in the DD case, we explicitly consider that the discrete fields $(\mathbf{p}_h, \phi_h)$ are also the solution to the variational formulation of the plain-case (6.2.11). Let us begin by a technical result.

Lemma 6.6.2 (`[A2, Lemma 6.4]`)**.** *Let the assumptions of lemma 6.6.1 hold. Let* $\mathbf{q} \in \mathbf{H}(\mathrm{div}, \mathcal{R}) \cap \mathbf{H}^\mu(\mathcal{R})$ *with* $0 < \mu < \frac{1}{2}$*, and define* $\delta\mathbf{q}_{fc} \in \mathbf{Q}_{f,h}$ *by* $\delta\mathbf{q}_{fc} \cdot \mathbf{n}_{|\Gamma_{fc}} = (\tilde{\mathbf{q}}_{c,R} \cdot \mathbf{n} - \tilde{\mathbf{q}}_{f,R} \cdot \mathbf{n})_{|\Gamma_{fc}}$ *and zero extension in* $\widetilde{\mathcal{R}_f} \setminus \Gamma_{fc}$*. It holds*

$$\|\delta\mathbf{q}_{fc}\|_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_f)} \lesssim h^\mu \left(\|\mathbf{q}_f\|_{\mathbf{H}^\mu(\widetilde{\mathcal{R}}_f)} + \|\operatorname{div} \mathbf{q}_f\|_{L^2(\widetilde{\mathcal{R}}_f)} \right).$$

Theorem 6.6.4 (`[A2, Theorem 6.5]`)**.** *Under the assumptions of Theorem 6.6.3 with* $r_{\max} < 1/2$*, one has for nested meshes:*

$$\forall \mu \in]0, r_{\max}[,\ \forall S_f \in H^\mu(\mathcal{R}), \quad \|\phi - \phi_h\|_{L^2(\mathcal{R})} \lesssim h^{2\mu} \|S_f\|_{H^\mu(\mathcal{R})}. \tag{6.6.24}$$

Proof. *Matching meshes.* In this case, one can use the theory already developed in §6.3.2 for the plain case, to conclude that (6.6.24) holds.
Nested meshes. The difficulty for non-matching meshes is that one can not define the global RTN-interpolant of $\mathbf{p}$ directly. Instead it is defined via its subdomain interpolants $(\widetilde{\mathbf{p}}_{i,R})_{i \in \tilde{\mathcal{I}}_\mathcal{R}}$. Introduce, for $i \in \tilde{\mathcal{I}}_\mathcal{R}$, $\mathcal{I}_i$ as the set of indices j such that $T_{j,h|\Gamma_{ij}} \subset T_{i,h|\Gamma_{ij}}$ (since we are dealing with nested meshes, it holds $T_{j,h|\Gamma_{ij}} \subset T_{i,h|\Gamma_{ij}}$ or $T_{i,h|\Gamma_{ij}} \subset T_{j,h|\Gamma_{ij}}$). We proceed as follows to obtain an $\mathbf{H}(\mathrm{div}, \mathcal{R})$-conforming approximant, i.e. an element of $\mathbf{Q}_h$. On all interfaces Γ_{ij}, introduce $\delta\mathbf{p}_{ij} \cdot \mathbf{n} = \widetilde{\mathbf{p}}_{c,R} \cdot \mathbf{n}_{|\Gamma_{ij}} - \widetilde{\mathbf{p}}_{f,R} \cdot \mathbf{n}_{|\Gamma_{ij}}$ where $\widetilde{\mathbf{p}}_{f,R}$ is the interpolant from the finer discretization on Γ_{ij}, resp. $\widetilde{\mathbf{p}}_{c,R}$ is the interpolant from the coarser discretization on Γ_{ij}. By construction, $\delta\mathbf{p}_{ij} \cdot \mathbf{n} = 0$ when $T_{i,h|\Gamma_{ij}} = T_{j,h|\Gamma_{ij}}$. Then $\delta\mathbf{p}_{ij} \cdot \mathbf{n}$ is extended by zero in $\widetilde{\mathcal{R}}_i$ to define an element of $\mathbf{Q}_{i,h}$; with a slight abuse of notation, we still denote the extension by $\delta\mathbf{p}_{ij}$. The $\mathbf{H}(\mathrm{div}, \mathcal{R})$-conforming approximant $\mathbf{p}_R \in \mathbf{Q}_h$ is then defined subdomain by subdomain as

$$\mathbf{p}_{i,R} = \widetilde{\mathbf{p}}_{i,R} + \sum_{j \in \mathcal{I}_i} \delta\mathbf{p}_{ij} \quad \text{for } i \in \tilde{\mathcal{I}}_\mathcal{R}.$$

Indeed, $[\mathbf{p}_R \cdot \mathbf{n}]_{\Gamma_{ij}} = 0$ for $(i,j) \in \mathcal{I}\mathcal{J}$ by direct inspection. It remains to evaluate

$$\|\mathbf{p} - \mathbf{p}_R\|^2_{\mathbf{H}(\mathrm{div}, \mathcal{R})} = \sum_{i \in \tilde{\mathcal{I}}_\mathcal{R}} \|\mathbf{p}_i - \mathbf{p}_{i,R}\|^2_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)}, \text{ with}$$
$$\|\mathbf{p}_i - \mathbf{p}_{i,R}\|_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)} \le \|\mathbf{p}_i - \widetilde{\mathbf{p}}_{i,R}\|_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)} + \sum_{j \in \mathcal{I}_i} \|\delta\mathbf{p}_{ij}\|_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)} \text{ for } i \in \tilde{\mathcal{I}}_\mathcal{R}.$$

Above, the fact that the index j belongs to $\mathcal{I}_i$ implies that if $\delta\mathbf{p}_{ij} \neq 0$, then the finer discretization on Γ_{ij} automatically originates from $\widetilde{\mathcal{R}}_i$. To evaluate $\|\delta\mathbf{p}_{ij}\|_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)}$, one uses the results of lemma 6.6.2 to find

$$\|\delta\mathbf{p}_{ij}\|_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)} \lesssim h^\mu \left(\|\mathbf{p}_i\|_{\mathbf{H}^\mu(\widetilde{\mathcal{R}}_i)} + \|\operatorname{div} \mathbf{p}_i\|_{L^2(\widetilde{\mathcal{R}}_i)} \right).$$

Again, this bound holds under the condition that the meshes are *quasi-uniform* on the part of the interface where they are non-matching. Due to (6.2.10), one has $\|\mathbf{p}_i - \mathbf{p}_{i,R}\|_{\mathbf{H}(\mathrm{div}, \widetilde{\mathcal{R}}_i)} \lesssim h^\mu \|S_f\|_{H^\mu(\mathcal{R})}$ for $i \in \tilde{\mathcal{I}}_\mathcal{R}$, and it follows that $\|\mathbf{p} - \mathbf{p}_R\|_{\mathbf{H}(\mathrm{div}, \mathcal{R})} \lesssim h^\mu \|S_f\|_{H^\mu(\mathcal{R})}$. As a consequence (following the proof of Theorem 6.2.15) we conclude that the estimate (6.6.24) holds. ⋄

6.6.5 Estimates for the generalized eigenproblem

Let $0 \le \mu < r_{\max}$ be given, we introduce an operator B_μ associated to the source problem (6.6.7): given $f \in H^\mu(\mathcal{R})$, we call $B_\mu f = \phi \in H^1(\mathcal{R})$ the second component of the triple $(\mathbf{p}, \phi, \phi_S)$ that solves the source problem with $S_f = \underline{\nu\Sigma}_f f$. For the same reason as in the plain case §6.3, B_μ is a bounded and compact operator.

Convergence in operator norm

Next, let us consider the discrete operator B_μ^h associated to the discrete source problem: given $f \in H^\mu(\mathcal{R})$, we call $B_\mu^h f$ the second component of the triple $(\mathbf{p}_h, \phi_h, \phi_{S,h})$ that solves (6.6.16) with source $S_f = \underline{\nu\Sigma}_f f$. Using estimate (6.6.24), we obtain, like in the plain case, the result below.

Theorem 6.6.5. *Under the assumptions of Theorem 6.6.3 with $r_{\max} < 1/2$ plus $\underline{\nu\Sigma}_f \in \mathcal{P}W^{1,\infty}(\mathcal{R})$, let $\mu \in]0, r_{\max}[$. Provided that the families of triangulations are* regular$^+$ *on every subdomain, one has for nested meshes:*

$$\|B_\mu - B_\mu^h\|_{\mathcal{L}(H^\mu(\mathcal{R}))} \lesssim h^{\widetilde{\theta}\mu}, \tag{6.6.25}$$

where $\widetilde{\theta} = \min_{i \in \tilde{\mathcal{I}}_\mathcal{R}} \theta_i > 0$, and for $i \in \tilde{\mathcal{I}}_\mathcal{R}$, θ_i is defined by (6.3.3) on $\widetilde{\mathcal{R}}_i$.

We conclude to the absence of spectral pollution.

Optimal convergence rate

Let the assumptions of Theorems 6.6.3 and 6.6.5 hold, and in particular the conditions for nested, *non-matching* meshes. We use the same notations as in §6.3.2. In particular, let $\widetilde{\omega}_\nu > 0$ be the regularity exponent associated to ν with respect to $(\widetilde{\mathcal{P}}H^{1+s}(\mathcal{R}))_{s>0}$, and introduce $\widetilde{\omega} = \min(\widetilde{\omega}_\nu, k+1)$.

Let $\mu \in [0, r_{\max}[$ be given. As we defined B_μ (resp. B_μ^h), we define A_μ and C_μ (resp. A_μ^h and C_μ^h): for $f \in H^\mu(\mathcal{R})$, we call $A_\mu f = \mathbf{p} \in \widetilde{\mathbf{Q}}$ and $C_\mu f = \phi_S \in M$ (resp. $A_\mu^h f = \mathbf{p}_h \in \widetilde{\mathbf{Q}}_h$ and $C_\mu^h f = \phi_{S,h} \in M_h$) the first and the third components of the triple $(\mathbf{p}, \phi, \phi_S)$ (resp. $(\mathbf{p}_h, \phi_h, \phi_{S,h})$) that solves (6.6.7) (resp. (6.6.16)) with source $S_f = \underline{\nu\Sigma}_f f$.

For the DD+L^2-jumps method, the transposition of Proposition 6.3.1 writes:

Proposition 6.6.2 (`[A2, Proposition 6.8]`). *For every φ in W, the following inequalities hold for the DD+L^2-jumps method with nested meshes:*

$$\|(B_\mu - B_\mu^h)\varphi\|_{L^2(\mathcal{R})} \lesssim h^{\widetilde{\omega}} \|\varphi\|_W; \quad \textit{and} \quad \|(A_\mu - A_\mu^h)\varphi\|_{\mathbf{H}(\mathrm{div},\mathcal{R})} \lesssim h^{\widetilde{\omega}} \|\varphi\|_W.$$

Estimate (6.3.6) on the gap between W and W_h is still valid: $\hat{\delta}(W, W_h) \lesssim h^{\widetilde{\omega}}$. Let E_h be the operator defined in (6.3.7). We recall that E_h and B_μ^h commute (lemma 6.3.2 holds). The restriction of E_h to W, denoted by F_h is a bijection that satisfies estimate (6.3.8), for h small enough. We will also make use of $\mathcal{S}_h = F_h^{-1} E_h - I$ that satisfies lemma 6.3.4 and Proposition 6.3.2. We recall that $\hat{B}_\mu = B_\mu|_W$ and $\hat{B}_\mu^h = F_h^{-1} B_\mu^h F_h$. The transposition of Theorem 6.3.2 is stated next. The proof is identical (replace ω by $\widetilde{\omega}$).

Theorem 6.6.6. *For h small enough, one has for the DD+L^2-jumps method with nested meshes:*

$$\|\hat{B}_\mu - \hat{B}_\mu^h\|_{\mathcal{L}(W)} \lesssim h^{2\widetilde{\omega}}. \tag{6.6.26}$$

Corollary 6.6.2. *For h small enough, the error on the eigenvalue for the DD+L^2-jumps method with nested meshes is given by: $|\nu - \nu_h| \lesssim h^{2\widetilde{\omega}}$.*

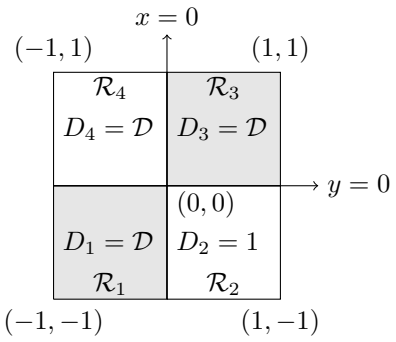
6.6.6 Numerical illustrations

The tests are carried out in two dimensions: the cartesian coordinates are denoted by (x, y). We define the discrete space of Lagrange multipliers M_h as in (6.6.21).

Checker board benchmark

We study a singular toy problem described on Dauge's website [125] for a magnetic problem and adapted here for the neutron diffusion equation with Neuman boundary condition. Set $\mathcal{R} :=]-1, 1[^2$, and

divide it into four subsquares (see Figure 6.1 left). Let D, be a scalar, piecewise-constant, coefficient:



h_c	h_c	h_c	h_c
h_c	h_f	h_f	h_c
h_c	h_f	h_f	h_c
h_c	h_c	h_c	h_c

Figure 6.1: The domain of study, and the subdomain mesh sizes.

$D := \mathcal{D} = 0.1$ in $\mathcal{R}_1 \cup \mathcal{R}_3$, and 1 elsewhere, $\Sigma_a = 1$ and $\underline{\nu\Sigma}_f = 1$. We consider the following problem:

$$\text{div}\, D\, \nabla \phi + \phi = \lambda \phi \text{ in } \mathcal{R}, \quad \nabla \phi \cdot \mathbf{n}_{|\partial\mathcal{R}} = 0 \text{ on } \partial\mathcal{R}. \tag{6.6.27}$$

The singularity exponent is $r_{\max} \approx 0.39$. Implementation is in Octave.

We study the error on the four first eigenvalues (excluding $\lambda_0 = 1$), with two different partitions $\{\widetilde{\mathcal{R}}_i\}_{1 \in \mathcal{I}}$. The results are given in Table 6.2, which data are:

- h: the meshsize,
- N_ϕ: the number of degrees of freedom of ϕ,
- $\varepsilon_{\lambda_i} = |\lambda_{h,i} - \lambda_i|/|\lambda_i|$: the relative error for the eigenvalue λ_i, $i = 1, 4$.

In the last line, we report the average rate of convergence of the computations.
The partition is based on $\widetilde{N} = 16$ square subdomains, represented in Figure 6.1 right. As indicated on this figure, the four centered subdomains have a meshsize equal to h_f whereas the other subdomains have a meshsize equal to $h_c = 2h_f$, so that the parameter is $h = h_c$. The results are given in Table 6.2.

$1/h$	N_ϕ	ε_{λ_1}	ε_{λ_2}	ε_{λ_3}	ε_{λ_4}
4	448	2.88×10^{-3}	3.92×10^{-2}	5.49×10^{-3}	2.00×10^{-2}
8	1 792	7.22×10^{-4}	2.36×10^{-2}	1.38×10^{-3}	5.00×10^{-3}
12	4 032	3.22×10^{-4}	1.74×10^{-2}	6.12×10^{-4}	2.22×10^{-3}
16	7 168	1.81×10^{-4}	1.40×10^{-2}	3.44×10^{-4}	1.25×10^{-3}
20	11 200	1.16×10^{-4}	1.18×10^{-2}	2.20×10^{-4}	8.00×10^{-4}
24	16 128	8.05×10^{-5}	1.02×10^{-2}	1.53×10^{-4}	5.05×10^{-4}
rate		2	0.76	2	2

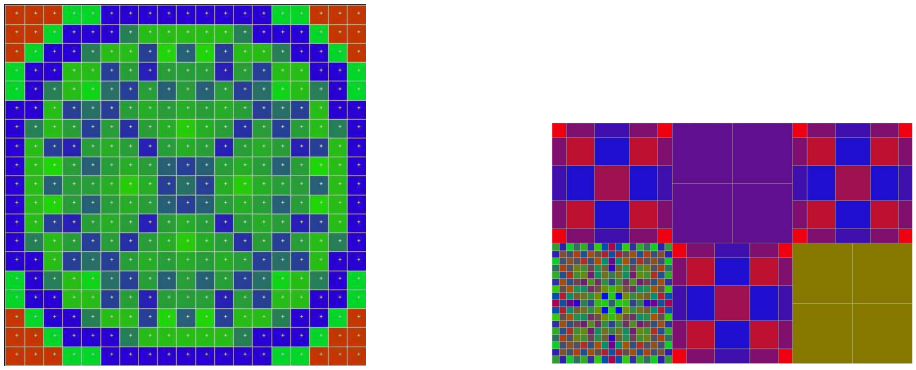
Table 6.2: Results with 16 subdomains.

Industrial simulation

This industrial test models a pressurized water large reactor core with heavy-steel reflector similar to the one described in [225]. The neutron transport equation is discretized using the multigroup simplified P_N (SP_N) equations, with two groups of energy, and SP_1 and SP_3 angular orders. We recall that, for each group, the neutron SP_1 equation is similar to the neutron diffusion equation, whereas the neutron SP_3 equation corresponds to two coupled neutron diffusion equations. The different homogenization steps

that allow to obtain the coefficients of this discretization on square cells lead to 229 different media. The coefficients are thus parametrized according to the medium, the energy group and the angular order, which depend respectively on the position, the energy and the direction of the neutrons.

In Figure 6.2(a), we draw a top view of a PWR-like core model, which has a Cartesian geometry. Each square, which represents a part of the reflector or an assembly, is made itself of cells, which are rectangular cuboids of $\mathbb{R}^3$. In Figure 6.2(b), we make a zoom on a patch of six (3×2) assemblies: each coloured square represents a cell containing fuel, absorbing or reflector material. In our model, the coefficients are polynomial (possibly constant) in each cell [143, A10, A11]. The global domain of the reactor core (see again Figure 6.2(a)) is represented by a rectangular cuboid of $\mathbb{R}^3$. In practice the coefficients characterizing the materials may differ from one cell to another by a factor of order 10 or more.



(a) Reactor core model

(b) Patch of six assemblies

Figure 6.2: $2D$ depiction of a PWR core and a zoom on six assemblies.

The subdomains $\{\widetilde{\mathcal{R}}_i\}_{1\leq i\leq 361}$ of the partition correspond to the 19×19 cells.In each subdomain, the coarser triangulation is also such that the coefficients are piecewise constant. The meshes of the subdomains are nested.

In neutronics, the quantity of interest is the inverse of the smallest eigenvalue, which is called the criticality, and is denoted by k_{eff}. Below, we make comparisons on the criticality, the reference value, denoted by k_{eff}^{ref}, being computed on a conforming mesh made of $1.5\,e+7$ (resp. $7.5 \times 10^{+6}$) rectangles in SP_1 (resp. SP_3).

In Table 6.3, we present the results obtained with the MINOS solver for different levels of refinement. The data are:

- h: the meshsize,
- N_ϕ: the spatial number of degrees of freedom of the neutron flux ϕ,
- ε_1 (resp. ε_3): the relative error made on the criticality $|k_{eff} - k_{eff}^{ref}|/k_{eff}^{ref}$, for a computation using the SP_1 (resp. SP_3) approximation.
- rate: the averaged rate of convergence.

Convergence rates are higher than 1, seemingly indicating the absence of strong singularities in the first eigenfunction. Instead, we hypothetize that we are still in the pre-asymptotic regime (for the first eigenfunction): on the one hand, the norm of the "more singular" part is small compared to the norm of the "more regular" part, and on the other hand there are only a few degrees of freedom per characteristic length (cf. Figure 6.2(b)). Further results are given in [P5, 163].

We proposed and analysed two domain decomposition methods. First a domain decomposition method with Optimized Robin interface conditions, which enables the use of non-overlapping subdomains

$1/h$	N_ϕ	ε_1	ε_3
285	$5.40 \times 10^{+5}$	1.35×10^{-4}	1.37×10^{-4}
380	$9.60 \times 10^{+5}$	8.01×10^{-5}	8.79×10^{-5}
570	$2.16 \times 10^{+6}$	4.10×10^{-5}	5.12×10^{-5}
665	$2.94 \times 10^{+6}$	3.26×10^{-5}	4.30×10^{-5}
950	$6.00 \times 10^{+6}$	2.09×10^{-5}	3.15×10^{-5}
rate		$h^{1.55}$	$h^{1.22}$

Table 6.3: Results with 361 subdomains.

and a speed up in the convergence. Second the DD+L^2-jumps method which allows non-conforming grids. These methods could be combined to leverage their strengths: a better conditionning in the global linear system and the use of non-conforming meshes. This approach is developped in [1] for a finite volume discretization.

In the next section, we study the numerical analysis of the neutron multigroup SP_N equation discretized with a discontinuous Galerkin finite element method.

6.7 Numerical analysis of the multigroup SP_N equation [A12]

Let $E \subset \mathcal{R}$. Let $W(E)$ be a scalar functional space on E. In this section, we set: $\underline{W}(E) = (W(E))^{\hat{N}\times G}$, $\underline{\mathbf{W}}(E) = \underline{W}(E)^d$, and we have for all $(\phi, \psi) \in \underline{W}(E) \times \underline{W}(E)$:

$$(\phi, \psi)_{\underline{W}(E)} = \sum_{g\in\mathcal{I}_G} \sum_{m\in\mathcal{I}_e} (\phi_m^g, \psi_m^g)_{W(E)} \text{ and } \|\phi\|^2_{\underline{W}(E)} = \sum_{g\in\mathcal{I}_G} \sum_{m\in\mathcal{I}_e} \|\phi_m^g\|^2_{W(E)}.$$

In the same way, for all $(\mathbf{u}, \mathbf{v}) \in \underline{\mathbf{W}}(E) \times \underline{\mathbf{W}}(E)$:

$$(\mathbf{u}, \mathbf{v})_{\underline{\mathbf{W}}(E)} = \sum_{g\in\mathcal{I}_G} \sum_{m\in\mathcal{I}_e} (\mathbf{u}_m^g, \mathbf{v}_m^g)_{\mathbf{W}(E)} \text{ and } \|\mathbf{u}\|^2_{\underline{\mathbf{W}}(E)} = \sum_{g\in\mathcal{I}_G} \sum_{m\in\mathcal{I}_e} \|\mathbf{u}_m^g\|^2_{\mathbf{W}(E)}.$$

We use the notation: $V(\mathcal{R}) := H_0^1(\mathcal{R})$.

Under assumptions of properties (6.1.10) and (6.1.11) the matrices $\mathbb{T}_e$ and $\mathbb{T}_o^{-1}$ are positive definite. Moreover, one can show that $\|\mathbb{H}\,\nabla\,\phi\|_{\underline{\mathbf{L}}^2(\mathcal{R})} \gtrsim \|\,\nabla\,\phi\|_{\underline{\mathbf{L}}^2(\mathcal{R})}$ [163]. We infer that the matrix $\mathbb{D}$ is positive definite and that there exists a constant $C_{\mathbb{D}} > 0$ such that for all $\xi \in \mathbb{R}^{\hat{N}\times G}$,

$$\|\mathbb{D}\,\xi\|^2 \leq C_{\mathbb{D}} \|\xi\|^2. \tag{6.7.1}$$

From now on, we suppose that this property holds.

6.7.1 Variational formulations

Consider Problem (6.1.8) posed with zero-flux boundary condition, i.e. for all $m \in \mathcal{I}_e$, $g \in \mathcal{I}_G$, $(\phi_m^g)_{|\partial\mathcal{R}} = 0$. The variational formulation reads:

$$\text{Find } \phi \in \underline{V}(\mathcal{R}) \text{ such that for all } \psi \in \underline{V}(\mathcal{R}) : c(\phi, \psi) = \ell(\psi), \tag{6.7.2}$$

where:

$$\left\{ \begin{array}{rcl} c : \underline{V}(\mathcal{R}) \times \underline{V}(\mathcal{R}) & \to & \mathbb{R} \\ c(\phi, \psi) & \mapsto & (\mathbb{D}\,\nabla\,\phi, \nabla\,\psi)_{\underline{\mathbf{L}}^2(\mathcal{R})} + (\mathbb{T}_e\phi, \psi)_{\underline{L}^2(\mathcal{R})} \end{array} \right. , \quad \left\{ \begin{array}{rcl} \ell : \underline{V}(\mathcal{R}) & \to & \mathbb{R} \\ \ell(\psi) & \mapsto & (S_f, \psi)_{\underline{L}^2(\mathcal{R})} \end{array} \right. .$$

The bilinear form $c(\cdot,\cdot)$ and the linear form $\ell(\cdot)$ are continuous. Under the hypothesis on $\mathbb{D}$, the bilinear form $c(\cdot,\cdot)$ is coercive: we can apply Lax–Milgram Theorem to state

Theorem 6.7.1. *Suppose that $\mathbb{D}$ is positive definite. For a given source term $S_f \in \underline{L}^2(\mathcal{R})$, it exists a unique $\phi \in \underline{V}(\mathcal{R})$ that solves Problem* (6.7.2)*. In addition,$\|\phi\|_{\underline{V}(\mathcal{R})} \lesssim \|S_f\|_{\underline{L}^2(\mathcal{R})}$.*

The variational formulation of Problem (6.1.9) reads:

$$\text{Find } \phi \in \underline{V}(\mathcal{R}) \text{ such that for all } \psi \in \underline{V}(\mathcal{R}) : \; c(\phi,\psi) = \lambda\, \ell_f(\phi,\psi), \tag{6.7.3}$$

where: $\left\{ \begin{array}{rcl} \ell_f : \underline{L}(\mathcal{R}) \times \underline{L}^2(\mathcal{R}) & \to & \mathbb{R} \\ \ell_f(\phi,\psi) & = & (\mathbb{M}_f\phi,\psi)_{\underline{L}^2(\mathcal{R})} \end{array} \right.$.
The bilinear form $\ell_f(\cdot,\cdot)$ is a continuous onto $\underline{L}(\mathcal{R}) \times \underline{V}(\mathcal{R})$. The space $\underline{V}(\mathcal{R})$ is a subset of $\underline{L}(\mathcal{R})$ with a compact embedding. We can then apply the theory of Babuška and Osborn in [18] to state

Theorem 6.7.2. *Suppose that $\mathbb{D}$ is positive definite. There exists a unique compact operator $T_f : \underline{L}^2(\mathcal{R}) \to \underline{L}^2(\mathcal{R})$ such that for all $\forall(\phi,\psi) \in \underline{L}^2(\mathcal{R}) \times \underline{V}(\mathcal{R})$: $c(T_f\phi,\psi) = \ell_f(\phi,\psi)$.*

Thus, the couple (ϕ,λ) is a solution to Problem 6.7.3 iff the couple (ϕ,λ^{-1}) is an eigenpair of operator T_f. Moreover, Problem (6.7.3) admits a countable number of eigenvalues.

6.7.2 Discretization with discontinuous Galerkin finite elements

Let $(\mathcal{T}_h)_h$ be a shape- and regular mesh sequence of $\mathcal{R}$ [130, def. 1.38]. We use the notations of §4.7, but the elements may be polygons when $d = 2$ or polyhedra when $d = 3$. We denote by N_∂ the maximum number of mesh faces composing the boundary of mesh elements

$$N_\partial := \max_{\ell \in \mathcal{I}_K} \text{card } \mathcal{I}_{F,\ell}. \tag{6.7.4}$$

Let us set $\underline{V}_h^k = (\mathcal{P}^k_{disc}(\mathcal{T}_h))^{\hat{N}\times G}$. We consider $c_h : \underline{V}_h^k \times \underline{V}_h^k \to \mathbb{R}$ such that

$$\forall(\phi_h,\psi_h) \in \underline{V}_h^k \times \underline{V}_h^k, \quad c_h(\phi_h,\psi_h) = c_{\mathcal{T}_h}(\phi_h,\psi_h) + c_{\mathcal{F}_h}(\phi_h,\psi_h), \tag{6.7.5}$$

with

$$\begin{array}{rcl} c_{\mathcal{T}_h}(\phi_h,\psi_h) & = & (\mathbb{D}\,\nabla_h\,\phi_h, \nabla_h\,\psi_h)_{\underline{\mathbf{L}}^2(\mathcal{R})} + (\mathbb{T}_e\phi_h,\psi_h)_{\underline{L}^2(\mathcal{R})}\,, \\ c_{\mathcal{F}_h}(\phi_h,\psi_h) & = & \displaystyle\sum_{f\in\mathcal{I}_F} \frac{\alpha}{h_f}\,([\phi_h]\,,[\psi_h])_{\underline{L}^2(F_f)} \\ & & \displaystyle - \sum_{f\in\mathcal{I}_F^i} \Big((\,\{\mathbb{D}\,\nabla_h\,\psi_h\}\cdot\mathbf{n}_f, [\phi_h])_{\underline{L}^2(F_f)} + (\,\{\mathbb{D}\,\nabla_h\,\phi_h\}\cdot\mathbf{n}_f, [\psi_h])_{\underline{L}^2(F_f)} \Big)\,, \end{array}$$

where α is a stabilization parameter.

The Symmetric Interior Penalty Galerkin method (SIPG) associated to Eq. (6.7.2) reads:

$$\text{Find } \phi_h \in \underline{V}_h^k \text{ such that } \forall\psi_h \in \underline{V}_h^k : \; c_h(\phi_h,\psi_h) = \ell(\psi_h). \tag{6.7.6}$$

Let us denote by $\mathcal{P}_h W^{2,p}(\mathcal{R}) = \{v \in L^p(\mathcal{R}) \,|\, \forall \ell \in \mathcal{I}_K,\, v_{|K_\ell} \in W^{2,p}(K_\ell)\}$.
We set $\mathcal{P}_h\underline{W}^{2,p}(\mathcal{R}) = \big(\mathcal{P}_h W^{2,p}(\mathcal{R})\big)^{\hat{N}\times G}$.

Hypothesis 6.7.1 (`Regularity of exact solution and space` $V^\star(\mathcal{R})$)**.** *Assume that $d \geq 2$ and that there is $2d/(d+2) < p \leq 2$ such that, for the exact solution $\phi \in \underline{V}^\star(\mathcal{R}) := \underline{V}(\mathcal{R}) \cap \mathcal{P}_h\underline{W}^{2,p}(\mathcal{R})$.*

This assumption requires $p > 1$ for $d = 2$ and $p > 6/5$ for $d = 3$. This holds for our assumptions on the coefficients, which are piecewise constant with respect to the triangulation [202]. In particular, we

observe that, in two space dimensions, $\phi \in \mathcal{P}_h\underline{W}^{2,p}(\mathcal{R})$ in polygonal domains. Moreover, using Sobolev embeddings [70, §9.3], this implies

$$\phi \in \underline{V}(\mathcal{R}) \cap \mathcal{P}_h\underline{\dot{H}}^{1+\alpha_p}(\mathcal{R}) \text{ where } \mathcal{P}_h\underline{\dot{H}}^{1+\alpha_p}(\mathcal{R}) := \bigcap_{0\le s<\alpha_p} \mathcal{P}_h\underline{H}^{1+s}(\mathcal{R}) \text{ with } \alpha_p = \frac{d+2}{2} - \frac{d}{p} > 0.$$

We state the following lemma [130, Lemma 1.46].[8]

Lemma 6.7.1. *Suppose that $(\mathcal{T}_h)_h$ is a shape- and contact-regular mesh sequence. Then there exists $C_{tr} > 0$, independent of h such that:*

$$\forall h > 0,\ \forall \psi_h \in \underline{V}_h^k,\ \forall \ell \in \mathcal{I}_K, \forall f \in \mathcal{I}_{F,\ell}, \quad h_\ell^{1/2}\|\psi_h\|_{\underline{L}^2(F_f)} \le C_{tr}\|\psi_h\|_{\underline{L}^2(K_\ell)}, \tag{6.7.7}$$

We aim at asserting the discrete coercivity using the following norm:

$$\forall \psi_h \in \underline{V}_h^k, \quad |||\psi_h|||_{sip}^2 := c_{\mathcal{T}_h}(\psi_h,\psi_h) + \|\psi_h\|_J^2 \text{ where } \|\psi_h\|_J^2 := \sum_{f\in\mathcal{I}_F} \frac{1}{h_f}\|[\psi_h]\|_{\underline{L}^2(F_f)}^2.$$

Under assumptions given in (6.1.7), there exists $\beta > 0$, independent of h such that for all $\psi_h \in \underline{V}_h^k$

$$c_{\mathcal{T}_h}(\psi_h,\psi_h) \ge \beta\left(\|\nabla_h\,\psi_h\|_{\underline{\mathbf{L}}^2(\mathcal{R})}^2 + \|\psi_h\|_{\underline{L}(\mathcal{R})}^2\right), \tag{6.7.8}$$

so that $|||\psi_h|||_{sip}^2 \ge \beta\left(\|\nabla_h\,\psi_h\|_{\underline{\mathbf{L}}^2(\mathcal{R})}^2 + \|\psi_h\|_{\underline{L}(\mathcal{R})}^2 + \|\psi_h\|_J^2\right)$. Following the proof of [130, Lemma 4.12], we can prove

Lemma 6.7.2 (`Discrete coercivity`)**.** *Let $\underline{\alpha} := (C_{tr})^2 N_\partial \frac{C_{\mathbb{D}}}{\beta}$, where C_{tr} depends on (6.7.7), N_∂ is defined in (6.7.4) and $C_{\mathbb{D}}$ is defined in (6.7.1). For all $\alpha \ge \underline{\alpha}$, the SIP bilinear form defined by (6.7.5) is coercive on $\underline{V}_h^k$ with respect to the $|||\cdot|||_{sip}$-norm, i.e.,*

$$c_h(\psi_h,\psi_h) \ge C_\alpha |||\psi_h|||_{sip}^2, \text{ with } C_\alpha := (\alpha - \underline{\alpha})\min\left(\frac{1}{2}, \beta\,(\alpha+\underline{\alpha})^{-1}\right).$$

Let $\underline{V}^{\star,h} = \underline{V}^\star(\mathcal{R}) + \underline{V}_h^k$ and the following norm

$$|||\psi|||_{sip,\star} := \left(|||\psi|||_{sip}^p + \sum_{\ell\in\mathcal{I}_K} (h_\ell)^{1+\gamma_p}\left\|\nabla\,\psi_\ell\cdot\mathbf{n}_{|\partial K_\ell}\right\|_{\underline{L}^p(\partial K_\ell)}\right)^{1/p},$$

where $\gamma_p = \frac{d(p-2)}{2}$ and $\psi_\ell = \psi_{|K_\ell}$. Following [130, §4.2], we obtain the following results.

Lemma 6.7.3 (`Boundedness`)**.** *There exists C_{bnd}, independent of h, such that for all $(\phi,\psi_h) \in \underline{V}^{\star,h}\times\underline{V}_h^k$*

$$c_h(\phi,\psi_h) \le C_{bnd}\,|||\phi|||_{sip,\star}|||\psi_h|||_{sip}$$

Theorem 6.7.3 (`Convergence`)**.** *Suppose that assumptions (6.1.7) hold. According to proposition 4.4.1, there exists $r_{\max}$ in $(0,1]$ such that $\phi \in \underline{V}(\mathcal{R}) \cap \mathcal{P}_h\underline{\dot{H}}^{1+r_{\max}}(\mathcal{R})$ where*

$$\mathcal{P}_h\underline{\dot{H}}^{1+r_{\max}}(\mathcal{R}) := \bigcap_{0\le\mu<r_{\max}} \mathcal{P}_h\underline{H}^{1+\mu}(\mathcal{R}).$$

Then, the solution of (6.7.6) is such that for all $\mu \in [0, r_{\max}[$:

$$\exists C > 0 \text{ independent of } h,\ |||\phi-\phi_h|||_{sip} \le C\,|\phi|_{\mathcal{P}_h\underline{W}^{2,p}(\mathcal{R})}\,h^\mu.$$

[8] In the case of simplicial meshes, it corresponds to Eq. (7.3.13).

We can then apply the Aubin-Nitsche trick similarly as in [130, Theorem 4.25].

Theorem 6.7.4 (L^2-norm estimate). *Suppose that assumptions* (6.1.7) *hold. According to propostion 4.4.1, there exists* $r_{\max}$ *in* $(0,1]$ *such that* $\phi \in \underline{V}(\mathcal{R}) \cap \underline{\dot{H}}^{1+r_{\max}}(\mathcal{R})$. *Then, the solution of* (6.7.6), ϕ_h *is such that for all* $\mu \in [0, r_{\max}[$: $\|\phi - \phi_h\|_{\underline{L}^2(\mathcal{R})} \lesssim h^{2\mu} \|S_f\|_{\underline{L}^2(\mathcal{R})}$.

We now have the tools to prove

Theorem 6.7.5 ([A12, Theorem 13]). *Let* μ *be the regularity of the eigenfunction* φ *associated to* λ, *and* $\omega = \min(\mu, k)$. *Let* λ_h *be the discrete eigenvalue associated to Problem* (6.7.3). *The following a priori error estimate holds:* $|\lambda - \lambda_h| \lesssim h^{2\omega}$.

Proof. We apply the theory developed in [6]. The proof is decomposed as follows. We first show that there is no spectral pollution. Then, we derive the error estimate. Let $\underline{V}^h := \underline{V}(\mathcal{R}) + \underline{V}_h^k$ and $\mathtt{E} : \underline{V}^h \to \underline{V}^h$ be the continuous spectral projector relative to λ defined by $\mathtt{E} = \frac{1}{2\pi\mathfrak{i}} \int_\Gamma \left(z - T_{|\underline{V}^h}\right)^{-1} dz$, where here Γ is a circle in the complex plane centred at λ which lies in $\rho(T_{|\underline{V}^h})$ and encloses no other points of $\sigma(T_{|\underline{V}^h})$. The absence of spectral pollution relies on two properties. First, using interpolation results [130, assumption 4.31] we have for all $\phi \in \mathtt{E}(\underline{V}^h)$, $\inf_{\psi_h \in \underline{V}_h^k} |||\phi - \psi_h|||_{sip} \lesssim h^\mu$. Second, using regularity results [202], we have for all $\phi_h \in \underline{V}_h^k$, for all $r \in [0, r_{\max}[$:

$$|||(T - T_h)\phi_h|||_{sip} \lesssim h^r \, |T\phi_h|_{\mathcal{P}_h \underline{W}^{2,p}(\mathcal{R})} \lesssim h^r \, \|T\phi_h\|_{\mathcal{P}_h \underline{\dot{H}}^{1+\alpha_p}(\mathcal{R})} \lesssim h^r \|\phi_h\|_{\underline{L}^2(\mathcal{R})} \lesssim h^r \|\phi_h\|_{\underline{L}^p(\mathcal{R})} \lesssim h^r |||\phi_h|||_{sip},$$

Applying [6, Theorem 3.7], we obtain that there is no spectral pollution. Moreover, we apply [6, Theorem 3.14] to state the error estimate on the eigenvalue, $|\lambda - \lambda_h| \lesssim \delta_h \delta_{*,h}$, where

$$\delta_h = \gamma_h + \left|\left|\left|(T - T_h)_{|\mathtt{E}(\underline{V}^h)}\right|\right|\right|_{sip} \text{ and } \delta_{*,h} = \gamma_{*,h} + \left|\left|\left|(T_* - T_{*,h})_{|\mathtt{E}_*(\underline{V}^h)}\right|\right|\right|_{sip},$$

with $\gamma_h = \delta(\mathtt{E}(\underline{V}^h), \underline{V}_h^k)$, $\gamma_{*,h} = \delta(\mathtt{E}_*(\underline{V}^h), \underline{V}_h^k)$, where $\delta(Y, Z) = \sup_{y \in Y, |||y|||_{sip}=1} \left(\inf_{z \in Z} |||y - z|||_{sip}\right)$, and $\mathtt{E}_* : \underline{V}^h \to \underline{V}^h$ is the continuous spectral projector of the adjoint operator $T_*|_{\underline{V}^h}$ relative to $\bar{\lambda}$. Using again elliptic regularity results [202] and Theorem 6.7.3, we obtain

$$\left|\left|\left|(T - T_h)_{|\mathtt{E}(\underline{V}^h)}\right|\right|\right|_{sip} \lesssim h^\mu \text{ and } \left|\left|\left|(T_* - T_{*,h})_{|\mathtt{E}_*(\underline{V}^h)}\right|\right|\right|_{sip} \lesssim h^\mu.$$

Using elliptic regularity results, we get $\|\varphi\|_{\mathcal{P}_h \underline{\dot{H}}^{1+\alpha_p}(\mathcal{R})} \lesssim \|\varphi\|_{\underline{L}^2(\mathcal{R})} \lesssim \|\varphi\|_{\underline{V}(\mathcal{R})}$. Applying Theorem 6.7.3, we infer that $\gamma_h \lesssim h^\mu$ and $\gamma_{*,h} \lesssim h^\mu$. ◇

In the next paragraph, we provide a numerical illustration.

6.7.3 Numerical illustration

We consider the test case Model 2, case 1 from the benchmark of Takeda and Ikeda [235]. The geometry of the core is three-dimensional and the domain is $\{(x, y, z) \in \mathbb{R}^3,\ 0 \le x \le 140 \text{ cm};\ 0 \le y \le 140\,\text{cm};\ 0 \le z \le 150 \text{ cm}\}$. This test is defined with $G = 4$ energy groups, isotropic scattering and vacuum boundary conditions. In Figure 6.3, we represent the cross-sectional geometry on the plane $z = 75\,\text{cm}$.

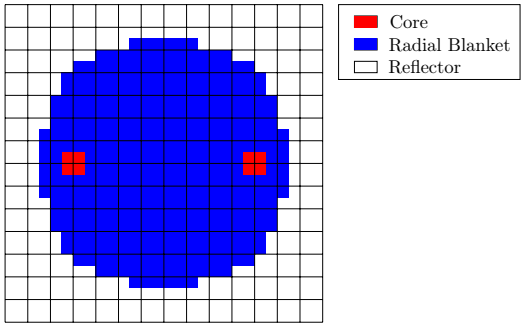


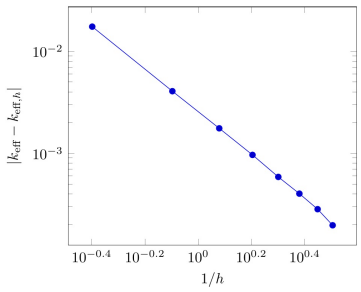
Figure 6.3: Cross-sectional view of the core ($z = 75$ *cm*).

Since the scattering is isotropic, the SP_3 formulation can easily be reformulated as a multigroup diffusion problem with 8 energy groups and an isotropic albedo boundary condition [29]. We then made

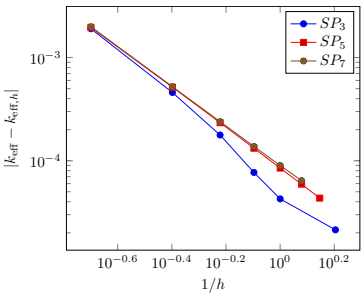
the computations with the PRIAM solver from the code CRONOS2 [185] for the conforming case and with the MINARET solver [198] from the APOLLO3® code [P10] for the SIPG discretization.

In Figure 6.4(a), we consider the convergence of the fundamental mode where we used the SP_3 formulation with Q_1 finite elements and a regular cartesian mesh of size h. The computations were performed using the CRONOS2 code. The approximated order of convergence is 2.22.

In Figure 6.4(b), we consider the convergence of the fundamental mode for different SP_N formulations with $\mathcal{P}^1_{disc}$ finite elements and a prismatic mesh of size h . The computations were performed using the APOLLO3® code. The approximated order of convergence is equal to 1.88 (resp. 1.96, 1.92) for the SP_3 (resp. SP_3, SP_5) approximation: it seems that the fundamental eigenmode is regular.



(a) Error on $k_{eff,h}$, Q^1.



(b) Error on $k_{eff,h}$, $\mathcal{P}^1_{disc}$.

Figure 6.4: Error on $k_{eff,h}$ with Q_1 or $\mathcal{P}^1_{disc}$ finite elements.

6.8 Coupling neutronics and thermal-hydraulics [A7, P2]

In this Section, we are interested in the coupling of the neutron diffusion equation and the low Mach number thermal-hydraulics model. This coupling is important to simulate nuclear reactors. Usually, the resolution is iterative. Here, we prove for a simple model that there exists a solution to the coupling and we propose a numerical procedure to compute it. This work could help to validate other resolution methods.

6.8.1 The model

We construct an analytic solution of the low Mach number thermal-hydraulics model

$$\left\{\begin{array}{rcll} \dfrac{d}{dz}(\rho\, u) & = & 0, & \text{(i)} \\ \dfrac{d}{dz}(\rho\, u^2 + \pi) & = & \rho\, g, & \text{(ii)} \\ \rho u \dfrac{dh}{dz} & = & \mathbb{E}\, \Sigma_f(h)\, \phi & \text{(iii)} \end{array}\right. \tag{6.8.1}$$

coupled to the simplified neutronic model based on the diffusion approximation with one energy group (cf. (6.1.14) for $d = 1$)

$$-\frac{d}{dz}\left[D(h)\frac{d\phi}{dz}\right] + \left[\Sigma_a(h) - \frac{\nu\Sigma_f(h)}{k_{eff}}\right]\phi = 0. \tag{6.8.2}$$

In (6.8.1) and (6.8.2), $z \in [0, L]$ is the spatial variable, $L > 0$ being the length of the nuclear core. Moreover, in (6.8.1), $\rho(z)$, $u(z)$, $\pi(z)$ and $h(z)$ are respectively the density, the velocity, the dynamical pressure and the internal enthalpy of the flow. The source term ρg is a volume force e.g the gravity field). The constant $\mathbb{E}$ is the energy released by a fission ($\mathbb{E} > 0$ is in Joule), $\Sigma_f(h)$ is the fission (macroscopic) cross section ($\Sigma_f(h) > 0$ is in m^{-1}) and $\phi(z)$ (the solution of (6.8.2) is the scalar neutron flux ($\phi(z) \geq 0$

is in $\mathrm{m}^{-2}\,.\,\mathrm{s}^{-1}$). Recall that in (6.8.2), $D(h)$ is the diffusion coefficient ($D(h) > 0$ is in m), $\Sigma_a(h)$ is the absorption (macroscopic) cross section ($\Sigma_a(h) > 0$ is in m^{-1}) and ν is the average number of neutron produced by a fission. Moreover, the density ρ and the internal enthalpy h are linked through the equation of state $\rho = \varrho(h)$ where $\varrho(\cdot)$ is a given function[9].

Using (6.8.1), we obtain $\rho\, u = D_e$ where $D_e > 0$ is a positive constant defining the flow rate. Thus, (6.8.1)-(iii) and (6.8.2) give the simplified neutronics/thermal-hydraulics system

$$\left\{\begin{array}{rcll} D_e \dfrac{dh}{dz} & = & \mathbb{E}\,\Sigma_f(h)\phi, & \text{(i)} \\ \\ -\dfrac{d}{dz}\left[D(h)\dfrac{d\phi}{dz}\right] + \left[\Sigma_a(h) - \dfrac{\nu\Sigma_f(h)}{k_{eff}}\right]\phi & = & 0. & \text{(ii)} \end{array}\right. \tag{6.8.3}$$

We supplement this system, written for $\phi \in H_0^1([0,L])$ and $h \in \mathcal{C}^1([0,L])$, with the constraint $\phi \geq 0$ on $[0,L]$ and with the boundary conditions

$$h(0) = h_e \quad \text{and} \quad h(L) = h_s. \tag{6.8.4}$$

Knowing $h(z)$, the density $\rho(z)$ is given by $\rho(z) = \varrho[h(z)]$. This allows to obtain the velocity $u(z)$ with $u(z) = \frac{D_e}{\rho(z)}$. At last, the dynamical pressure $\pi(z)$ is obtained by integrating (6.8.1)-(ii) and by using the boundary condition $\pi(L) = \pi_*$ where π_* is the pressure at the outlet of the nuclear core. Finally, one notice that thanks to the enthalpy equation (6.8.3)-(i) which imposes the relation $D_e(h_s - h_e) = \mathbb{E}\int_0^L \Sigma_f(h(z))\phi(z)dz$, System (6.8.3)–(6.8.4) reads

$$\frac{dh}{dz} = \frac{(h_s - h_e)\mathbb{E}\Sigma_f(h)\,\phi}{\displaystyle\int_0^L \mathbb{E}\Sigma_f(h(z))\phi(z)\mathrm{d}z} \text{ and } -\frac{d}{dz}\left[D(h)\frac{d\phi}{dz}\right] + \left[\Sigma_a(h) - \frac{\nu\Sigma_f(h)}{k_{eff}}\right]\phi = 0,$$

again coupled to the boundary conditions (6.8.4). This new system, which is often used in the field of neutronics/thermal-hydraulics coupling, is invariant with respect to the transformation $\phi \mapsto \mu\phi$ where μ is a positive constant.

The traditional approach for this type of coupling is to solve iteratively the two equations. In [128], the authors construct an analytical solution to (6.8.3)–(6.8.4) when $D(h)$ and $\Sigma_f(h)$ are positive constants, $\Sigma_a(h)$ being a non-constant function of h (to enforce the coupling). Here, we generalize this result by supposing that $D(h)$ and $\nu\Sigma_f(h)$ are also functions depending on h.

6.8.2 An analytical solution [A7]

To construct an analytical solution of (6.8.3)–(6.8.4), we assume that the given functions $\Sigma_f(h)$, $\Sigma_a(h)$ and $D(h)$ verify the two following hypotheses:

Hypothesis 6.8.1. *The enthalpy always belongs to a fixed domain $[h_{min}, h_{max}]$ on which $\Sigma_f(h)$, $\Sigma_a(h)$ and $D(h)$ are continuous functions.*

Hypothesis 6.8.2. *There exists $\alpha_f > 0$, $\alpha_a > 0$ and $\alpha_d > 0$ such that in $[h_{min}, h_{max}]$*

$$\Sigma_f(h) \geq \alpha_f, \quad \Sigma_a(h) \geq \alpha_a \quad \text{and} \quad D(h) \geq \alpha_d. \tag{6.8.5}$$

Using a classical property of Sturm-Liouville operators [110, Chapter 8], we have the following result[10]:

Lemma 6.8.1 (`[A7, Lemma 2.1]`). *Under Hypotheses 6.8.1 and 6.8.2, for any $(h_e, h_s) \in [h_{min}, h_{max}]^2$,*

[9] The fact that the equation of state $\varrho(h)$ depends only on h is a consequence of the low Mach speed [127].
[10] We can also use Sobolev embedding theorem.

there exists a unique function $\mathcal{X} \in \mathcal{C}^1([h_e, h_s])$ *solution of*

$$-\frac{d}{dh}\left[D(h)\nu\Sigma_f(h)\frac{d\mathcal{X}}{dh}\right] = \frac{\Sigma_a(h)}{\nu\Sigma_f(h)}, \quad \mathcal{X}(h_e) = \mathcal{X}(h_s) = 0 \tag{6.8.6}$$

and there exists a unique function $\mathcal{Y} \in \mathcal{C}^1([h_e, h_s])$ *solution of*

$$-\frac{d}{dh}\left[D(h)\nu\Sigma_f(h)\frac{d\mathcal{Y}}{dh}\right] = 1, \quad \mathcal{Y}(h_e) = \mathcal{Y}(h_s) = 0. \tag{6.8.7}$$

Moreover, $\mathcal{X}$ *and* $\mathcal{Y}$ *are positive on* $]h_e, h_s[$.

We will use these two functions to construct a direct solution to problem (6.8.3)–(6.8.4). Write $\phi(z) = (D_e \,/\, \mathbb{E}\Sigma_f(h))\, \dfrac{dh}{dz}$ and use it in (6.8.3)-(ii):

$$-\frac{d}{dz}\left[D(h)\frac{d\phi}{dz}\right] + \frac{\nu D_e}{\mathbb{E}}\left[\frac{\Sigma_a(h)}{\nu\Sigma_f(h)} - \frac{1}{k_{eff}}\right]\frac{dh}{dz} = 0.$$

Thus, using (6.8.6) and (6.8.7), we find

$$-\frac{d}{dz}\left[D(h)\frac{d\phi}{dz}\right] - \frac{\nu D_e}{\mathbb{E}}\left\{\frac{d}{dh}\left[D(h)\nu\Sigma_f(h)\frac{d\mathcal{X}}{dh}\right] - \frac{1}{k_{eff}}\frac{d}{dh}\left[D(h)\nu\Sigma_f(h)\frac{d\mathcal{Y}}{dh}\right]\right\}\frac{dh}{dz} = 0.$$

Using the chain rule, we recognize the (weak)[11] derivative with respect to z:

$$D(h)\nu\Sigma_f(h)\frac{d\mathcal{X}}{dh} - \frac{1}{k_{eff}}D(h)\nu\Sigma_f(h)\frac{d\mathcal{Y}}{dh}.$$

There exists a constant K_0 such that

$$-D(h)\frac{d\phi}{dz} - \frac{\nu D_e}{\mathbb{E}}\left[D(h)\nu\Sigma_f(h)\frac{d\mathcal{X}}{dh} - \frac{1}{k_{eff}}D(h)\nu\Sigma_f(h)\frac{d\mathcal{Y}}{dh} - K_0\right] = 0.$$

We multiply this equation by the continuous function $\phi(z)/D(h) = (D_e/D(h)\,\mathbb{E}\Sigma_f(h))\, \dfrac{dh}{dz}$:

$$-\phi(z)\frac{d\phi}{dz} - \left(\frac{\nu D_e}{\mathbb{E}}\right)^2\left[\frac{d\mathcal{X}}{dh} - \frac{1}{k_{eff}}\frac{d\mathcal{Y}}{dh} - \frac{K_0}{D(h)\nu\Sigma_f(h)}\right]\frac{dh}{dz} = 0.$$

Under hypothesis 6.8.1 and 6.8.2, we can define the $\mathcal{C}^1$-diffeomorphism

$$\theta : [h_e, h_s] \to [0, \theta_s] \text{ such that } \theta(h) = \int_{h_e}^{h} \left(D(h')\nu\Sigma_f(h')\right)^{-1} \mathrm{d}h', \quad \text{with} \quad \theta_s = \theta(h_s). \tag{6.8.8}$$

Using the chain rule and the weak derivative of ϕ^2, we get that it exists a constant K_1 such that

$$-\frac{1}{2}\phi^2(z) - \left(\frac{\nu D_e}{\mathbb{E}}\right)^2\left[\mathcal{X}(h) - \frac{1}{k_{eff}}\mathcal{Y}(h) - K_0\,\theta(h) - K_1\right] = 0.$$

The boundary conditions $\phi(0) = \phi(L) = 0$ and $\mathcal{X}(h_e) = \mathcal{X}(h_s) = \mathcal{Y}(h_e) = \mathcal{Y}(h_s) = 0$ yield $K_0 = K_1 = 0$. Finally, by replacing $\phi(z)$ by $(D_e/\mathbb{E}\Sigma_f(h))\, \dfrac{dh}{dz}$, we obtain

$$\frac{1}{2}\left(\frac{dh}{dz}\right)^2 = (\nu\Sigma_f(h))^2\left[\frac{1}{k_{eff}}\mathcal{Y}(h) - \mathcal{X}(h)\right].$$

[11] We have only here a weak derivative because, although $\mathcal{X}$ is of class $\mathcal{C}^1([h_e, h_s])$, $D(h)\nu\Sigma_f(h)$ is *a priori* only continuous.

A necessary condition on the solution h is thus $\frac{1}{k_{eff}}\mathcal{Y}(h(z)) - \mathcal{X}(h(z)) \geq 0$ on $[0, L]$ that is to say $k_{eff} \in]0, k_\infty]$ ($k_{eff} \neq 0$ since $L \neq 0$) where:

$$k_\infty := \min_{h\in[h_e,h_s]} \frac{\mathcal{Y}(h)}{\mathcal{X}(h)} = \frac{\mathcal{Y}(h_*)}{\mathcal{X}(h_*)}.$$

Let us now suppose that $\phi(z) \geq 0$. Thus, $h'(z) \geq 0$. When $k_{eff} \in]0, k_\infty[$, this allows to deduce from the previous relation that

$$\left(\nu\Sigma_f(h)\sqrt{2\left[\frac{1}{k_{eff}}\mathcal{Y}(h) - \mathcal{X}(h)\right]}\right)^{-1} \frac{dh}{dz} = 1. \tag{6.8.9}$$

Since $h(0) = h_e$ and $h(L) = h_s$, we deduce that:

$$\int_{h_e}^{h_s} \left(\nu\Sigma_f(h)\sqrt{2\left[\frac{\mathcal{Y}(h)}{k_{eff}} - \mathcal{X}(h)\right]}\right)^{-1} \mathrm{d}h = L. \tag{6.8.10}$$

Let us now define $\mathcal{L}(k)$ on $[0, k_\infty[$ so that:

$$\mathcal{L}(k) = \int_{h_e}^{h_s} \left(\nu\Sigma_f(h)\sqrt{2\left[\frac{\mathcal{Y}(h)}{k} - \mathcal{X}(h)\right]}\right)^{-1} \mathrm{d}h. \tag{6.8.11}$$

We cannot write that $(\mathcal{X}, \mathcal{Y}) \in \left(C^2([h_e, h_s])\right)^2$ since Σ_f and D are *a priori* only continuous on $[h_e, h_s]$. Nevertheless,with an *ad hoc* change of variable, one has extra regularity: letting $x(\theta) = \mathcal{X}(h(\theta))$, $y(\theta) = \mathcal{Y}(h(\theta))$, equations (6.8.6) and (6.8.7) become:

$$-\frac{d^2x}{d\theta^2} = D(h(\theta))\,\Sigma_a(h(\theta)) := f_x(\theta) \quad \text{and} \quad -\frac{d^2y}{d\theta^2} = D(h(\theta))\,\nu\Sigma_f(h(\theta)) := f_y(\theta).$$

These new functions satisfy the following regularity results:

Lemma 6.8.2 (`[A7, Lemma 2.2]`). *Under Hypotheses 6.8.1 and 6.8.2, the functions x, y belong to $\mathcal{C}^2([0, \theta_s])$. Moreover, the function $\frac{x}{y}$ belongs to $\mathcal{C}^1([0, \theta_s]) \cap \mathcal{C}^2(]0, \theta_s[)$. At last, $y(\theta) = (\theta_s - \theta)\theta\, r(\theta)$ with $r(\theta) \geq \frac{1}{2}\,\nu\,\alpha_f\,\alpha_d > 0$ and $r \in \mathcal{C}^0([0, \theta_s])$.*

Proof. Due to Hyp. 6.8.2, the functions x and y are concave on $[0, \theta_s]$, positive on $]0, L[$ and $x'(0) > 0$, $y'(0) > 0$, $x'(\theta_s) < 0$, $y'(\theta_s) < 0$. Due to Hyp.6.8.1, the functions f_x and f_y are continuous. Morevoer:

$$x(\theta) = -(\theta_s - \theta)\int_0^1 x'[\theta_s + t(\theta - \theta_s)]\,\mathrm{d}t, \quad y(\theta) = -(\theta_s - \theta)\int_0^1 y'[\theta_s + t(\theta - \theta_s)]\,\mathrm{d}t, \tag{6.8.12}$$

$$y(\theta) = -(\theta_s - \theta)\,\theta\, r(\theta) \quad \text{where} \quad r(\theta) = \int_0^1\int_0^1 t\, f_y\,[t\theta_s + \tau\, t(\theta - \theta_s)]\,\mathrm{d}\tau\,\mathrm{d}t. \tag{6.8.13}$$

Since $y'(\theta_s) < 0$, one deduces from (6.8.12) that $\frac{x}{y} \in \mathcal{C}^1\left([\frac{\theta_s}{2}, \theta_s]\right)$. A similar equality shows that $\frac{x}{y} \in \mathcal{C}^1\left([0, \frac{\theta_s}{2}]\right)$. Note that $r \in \mathcal{C}^0([0, \theta_s])$ and, under Hyp. 6.8.2, we find that r is bounded below by $\frac{1}{2}\nu\alpha_f\alpha_d$. ⋄

Next, we will need the

Lemma 6.8.3 (`[A7, Lemma 2.3]`). *Under Hyp. 6.8.1 and 6.8.2, the function $\mathcal{L}(k)$ defined in* (6.8.11) *is an increasing one to one function that maps $[0, k_\infty[$ into $[0, +\infty[$.*

Proof. The proof comes from the two following observations:
(i) Recall that $\theta(h)$ is given by (6.8.8) and $\theta_s = \theta(h_s)$. Notice that:

$$\mathcal{L}(k) = \sqrt{\frac{k}{2}}\int_0^{\theta_s} (y(\theta) - k\,x(\theta))^{-\frac{1}{2}}\; D\,(h(\theta))\,\mathrm{d}\theta \leq \alpha_d\sqrt{\frac{k}{2}}\int_0^{\theta_s} (y(\theta) - k\,x(\theta))^{-\frac{1}{2}}\;\mathrm{d}\theta.$$

We deduce from Lemma 6.8.2 that $\int_0^{\theta_s}(y(\theta))^{-1/2}\mathrm{d}\theta < +\infty$.
Thus, $\lim_{k\to 0}\mathcal{L}(k) = 0$ and by continuity, $\mathcal{L}(0) = 0$.
(ii) According to Lemma 6.8.2, $\frac{x}{y}$ is continuous on $[0,\theta_s]$ and admits a maximum equal to $\frac{1}{k_\infty}$. There are two possibilities:
(ii-1) There exists $\theta(h_*) \in]0,\theta_s[$ such that $\frac{x}{y}[\theta(h_*)] = \frac{1}{k_\infty}$. Since $\frac{x}{y} \in \mathcal{C}^2(]0,\theta_s[)$, applying Taylor-Lagrange formula, it exists $\tilde{\theta}[\theta,\theta(h_*)] \in [\theta(h_*)-\delta,\theta(h_*)+\delta]$ such that $\frac{x(\theta)}{y(\theta)} = \frac{1}{k_\infty} + \frac{1}{2}[\theta-\theta(h_*)]^2\left(\frac{x}{y}\right)''(\tilde{\theta})$. In this case, there exist $\delta > 0$ and $m_0 > 0$ such that $[\theta(h_*)-\delta,\theta(h_*)+\delta] \subset]0,\theta_s[$ and

$$\forall\theta \in [\theta(h_*)-\delta,\theta(h_*)+\delta]: \quad \frac{1}{k}-\frac{1}{k_\infty} = \frac{1}{k}-\frac{x}{y}[\theta(h_*)] \le \frac{1}{k}-\frac{1}{k_\infty}+\frac{1}{2}m_0[\theta-\theta(h_*)]^2. \tag{6.8.14}$$

Finally, since we have also $\mathcal{L}(k) = \int_0^{\theta_s} D(h(\theta))\left(\sqrt{2y(\theta)\left[\frac{1}{k}-\frac{x(\theta)}{y(\theta)}\right]}\right)^{-1}\mathrm{d}\theta$, we obtain (using a lower bound on $(y(\theta))^{-\frac{1}{2}}$) that there exists $m > 0$ such that

$$\mathcal{L}(k) \ge m\int_{\theta(h_*)-\delta}^{\theta_*+\delta}\left(\sqrt{\frac{1}{k}-\frac{1}{k_\infty}+\frac{1}{2}m_0[\theta-\theta_*]^2}\right)^{-1}\mathrm{d}\theta =: J(k,\delta), \text{ where } \theta_* = \theta(h_*).$$

Changing the variable $\theta = \theta_* + \sqrt{\frac{2}{m_0}\left(\frac{1}{k}-\frac{1}{k_\infty}\right)}\,t$, we have $J(k,\delta) = 2m\sqrt{\frac{2}{m_0}}\int_0^{\beta(k,\delta)}(1+t^2)^{-1/2}\mathrm{d}t$ with $\lim_{k\to k_\infty}\beta(k,\delta) := \delta\left(\frac{2}{m_0}\left(\frac{1}{k}-\frac{1}{k_\infty}\right)\right)^{-1/2} = +\infty$. Hence, $\lim_{k\to k_\infty} J(k,\delta) = +\infty$ which implies that $\lim_{k\to k_\infty}\mathcal{L}(k) = +\infty$.
(ii.2) The maximum of $\frac{x}{y}(\theta)$ is obtained at θ_s.
In this case, one writes $\frac{x}{y}(\theta) = \frac{1}{k_\infty} + (\theta-\theta_s)\left(\frac{x}{y}\right)'[\tilde{\theta}(\theta,\theta_s)]$, with $\left(\frac{x}{y}\right)'[\tilde{\theta}(\theta,\theta_s)] \ge 0$ on $[\theta_s-\delta,\theta_s]$.
Hence, there exists m_0 such that $\frac{1}{k}-\frac{x}{y}(\theta) \ge \frac{1}{k}-\frac{1}{k_\infty}+m_0(\theta_s-\theta)$ i.e. $y(\theta)\left[\frac{1}{k}-\frac{x}{y}(\theta)\right] \ge (\theta_s-\theta)\theta r(\theta)\left[\frac{1}{k}-\frac{1}{k_\infty}+m_0(\theta_s-\theta)\right]$ (r is defined in Lemma 6.8.2). Thus, using a lower bound of $m_0\theta r(\theta)$ on $[\theta_s-\delta,\theta_s]$, we find $M > 0$ such that

$$\mathcal{L}(k) \ge M\int_{\theta_s-\delta}^{\theta_s}\left((\theta_s-\theta)\left[\frac{1}{m_0}\left(\frac{1}{k}-\frac{1}{k_\infty}\right)+\theta_s-\theta\right]\right)^{-1/2}\mathrm{d}\theta = M'\int_0^{t_0}(t+t^2)^{-1/2}\mathrm{d}t,$$

where M' is another positive constant and $t_0 = \frac{\delta m_0}{\frac{1}{k}-\frac{1}{k_\infty}}$. Hence, $\lim_{k\to+\infty}\mathcal{L}(k) = +\infty$.
The case where the maximum is reached at $\theta = 0$ is treated similarly. ◇

We can now propose a three steps resolution to solve (6.8.3)–(6.8.4):

Proposition 6.8.1 (`[A7, Proposition 2.1]`)**.** *Let us suppose that* (6.8.3)–(6.8.4) *admits a solution* (h,ϕ,k_{eff}) *belonging to* $\mathcal{C}^1([0,L])\times H_0^1([0,L])\times\mathbb{R}_+$ *and such that* $\phi(z) \ge 0$ *on* $[0,L]$*. Then:*

1. $k_{eff} \in]0,k_\infty]$*. Moreover,* k_{eff} *is solution of* $L = \mathcal{L}(k_{eff})$ *(cf. eq.* (6.8.11)*).*
2. *Once* k_{eff} *is obtained,* $h(z)$ *is solution of*

$$\int_{h_e}^{h(z)}\left(\nu\Sigma_f(h)\sqrt{2\left[\frac{\mathcal{Y}(h)}{k_{eff}}-\mathcal{X}(h)\right]}\right)^{-1}\mathrm{d}h = z. \tag{6.8.15}$$

3. *Once* $h(z)$ *is obtained,* $\phi(z)$ *is given by*

$$\phi(z) = \frac{\nu D_e}{\mathbb{E}}\sqrt{2\left[\frac{1}{k_{eff}}\mathcal{Y}(h)-\mathcal{X}(h)\right]}. \tag{6.8.16}$$

Lemma 6.8.2 and Proposition 6.8.1 allow to prove the main result on the existence of a solution to Problem (6.8.3)–(6.8.4):

Theorem 6.8.1 (`[A7, Theorem 2.1]`)**.** *Suppose that Hypotheses 6.8.1 and 6.8.2 hold. Then, there exists a unique solution* $(h,\phi,k_{eff}) \in C^1([0,L])\times H_0^1([0,L])\times]0,k_\infty[$ *to Equations* (6.8.3)–(6.8.4) *such that* $\phi(z) \ge 0$ *on* $[0,L]$*. Moreover,* $\phi(z) > 0$ *on* $]0,L[$ *and* $\phi \in \mathcal{C}^1([0,L])$*.*

Proof. Using Lemma 6.8.2, we obtain the existence and unicity of (h,ϕ,k_{eff}) constructed in Proposition 6.8.1 (with $k_{eff} \in]0,k_\infty[$ since $L > 0$ and $L < +\infty$). Using Hypothesis 6.8.1, relations (6.8.15) and (6.8.16) prove that $(h,\phi) \in \mathcal{C}^1([0,L])\times H_0^1([0,L])$. And since $k_{eff} \in]0,k_\infty[$, we have $\frac{1}{k_{eff}}\mathcal{Y}(h)-\mathcal{X}(h) > 0$ on $]h_e,h_s[$ i.e. $\phi(z) > 0$ on $]0,L[$. Moreover, (6.8.3)-(ii) implies that $D(h)\phi' \in \mathcal{C}^1([0,L])$. Thus, $\phi' \in \mathcal{C}^0([0,L])$, i.e. $\phi \in \mathcal{C}^1([0,L])$. ◇

The regularity of h (resp. ϕ) is *a priori* not better that $\mathcal{C}^1([0,L])$ because we only suppose that $\Sigma_f(h)$ (resp. $D(h)$) is continuous. We now increase the regularity of $\Sigma_f(h)$ by supposing that:

Hypothesis 6.8.3. *The function $\Sigma_f(h)$ belongs to $\mathcal{C}^1([h_{\min}, h_{\max}])$.*

This allows to obtain the following result:

Proposition 6.8.2 (`[A7, Proposition 3.1]`). *Assume Hypotheses 6.8.1-6.8.3 and let (h, k_{eff}) be the solution obtained in Theorem 6.8.1. Then:*
(i) The operator

$$P(h) := -(\nu\Sigma_f(h))^{-\frac{1}{2}} \frac{d}{dz}\left\{ D(h)\frac{d}{dz}\left[(\nu\Sigma_f(h))^{-\frac{1}{2}}\right]\right\} + \frac{\Sigma_a(h)}{\nu\Sigma_f(h)}$$

is a self-adjoint positive operator belonging to $\mathcal{L}[H_0^1([0,L]), H^{-1}([0,L])]$.
(ii) The real $\frac{1}{k_{eff}}$ is the smallest eigenvalue of $P(h)$.
(iii) The function $(\nu\Sigma_f(h))^{\frac{1}{2}}\phi$ where ϕ is given by (6.8.16) *is an eigenvector of $P(h)$ associated to the eigenvalue $\frac{1}{k_{eff}}$.*

We exhibited an analytic solution of a simplified stationary neutronics/thermal-hydraulics model with minimal hypotheses on the absorption and fission cross sections, and on the diffusion coefficient. The construction of this solution underlines in particular that the coupling is not an eigenvalue problem although it is possible to prove that, when the internal enthalpy is known, the scalar neutron flux is also solution of an eigenvalue problem. Nevertheless, since this coupling is non-linear, the internal enthalpy cannot be known independently of the scalar neutron flux. Even though one recovers the classical set-up of neutronic equations (where the flux is an eigenvector and $\frac{1}{k_{eff}}$ is the smallest eigenvalue), it is only a *a posteriori* result for the coupling problem

6.8.3 Numerical illustation [P2]

We apply the three steps resolution of proposition 6.8.1 to solve equations (6.8.3)–(6.8.4) with data $(\Sigma_a^i)_{i=1}^N$, $(\nu\Sigma_f^i)_{i=1}^N$, $(D^i)_{i=1}^N$ that are known for a given list of temperatures $(T_i)_{i=1}^N$, and we use the linear approximation $h = C_p\,T$, C_p being the thermal capacity. We assume that $h_e = C_p\,T_e$ and $h_s = C_p\,T_s$ are given. We perform the following steps that we summarize here.

First, we consider the polynomial $(\nu\Sigma_f(h))^{-1}$ of degree $N-1$ which value is $(\nu\Sigma_f^i)^{-1}$, at each point $(h_1,\ldots,h_N)$, with $h_i = C_p\,T_i$. We express all quantities using the normalized variable

$$X(h) = -1 + 2\frac{U(h)-U(h_e)}{U(h_s)-U(h_e)}, \text{ with } U(h) = \int_0^h (\nu\Sigma_f(h'))^{-1}dh'.$$

Second, we built polynomials $\tilde{\mathcal{X}}(h)$ and $\tilde{\mathcal{Y}}(h)$ which are approximations of $\mathcal{X}(h)$ and $\mathcal{Y}(h)$ with P^3 finite elements. We then obtain an approximation of problem (6.8.9). With an ad hoc transformation, we need to solve the following problem to obtain k_{eff}

$$F_{m(k_eff)}(\pi/2) = \sqrt{\frac{C_0(k_{eff})}{2}}L,$$

where $F_m(\vartheta) =: \displaystyle\int_0^{\vartheta} (1-m\sin^2\tau)^{-1/2}\mathrm{d}\tau$ is the Jacobi elliptic function, and C_0 depends on the coefficients of polynomials $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$. We solve this equation numerically. We find successively k_{eff}, $h(z)$ and $\phi(z)$.

We consider the following example: $L = 4.06\,\mathrm{m}$, $N=3$, $T_e = 559\,\mathrm{K}$, $T_s = 633\,K$. Using the values of the fission cross section $(\nu\Sigma_f^i)_{i=1}^3$, we obtain the dimensionless values of $X_i := X(T_i)$ from which one deduces

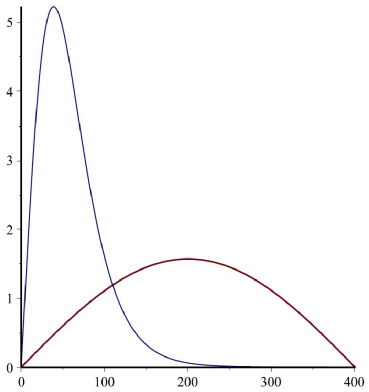
$$\mathcal{X}(h(X)) = (1-X^2)(a_X\,X + b_X\,Y)\,\mathcal{Y}(h(X)) = (1-X^2)(a_Y\,X + b_Y\,Y)$$

with $(a_X, b_X) = (-3.38 \times 10^{-3}, 9.91 \times 10^{-2}), (a_Y, b_Y) = (-4.68 \times 10^{-3}, 1.02 \times 10^{-1})$. We also compute the interpolation polynomials (in cm^{-1}, cm, cm)

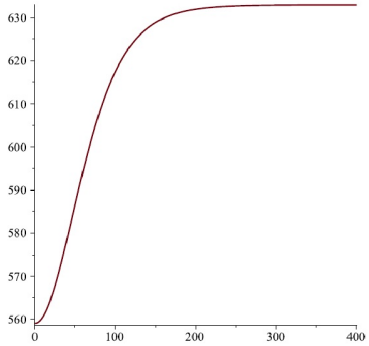
$$\begin{array}{lcl} D(X) & = & 2.49 \times 10^{-1}\,(1 + 1.656.10^{-2}X + 1.06.10^{-2}X^2), \\ \nu\Sigma_f(X) & = & 5.15 \times 10^{-2}\,(1 - 1.04.10^{-1}X - 3.17.10^{-2}X^2), \\ \Sigma_a(X) & = & 4.98 \times 10^{-2}\,(1 - 6.91.10^{-2}X - 1.48.10^{-2}X^2). \end{array}$$

We obtain: $k_{eff} = 1.014$. In Figure 6.5(a), we represent the neutron flux obtained with constant coefficients in red (a sinus), and the neutron flux solution of the coupling in blue. In Figure 6.5(b), we represent the corresponding temperature. We note that the neutron flux decreases (more neutrons are captured) as the temperature increases, which is known as the thermal Doppler broadening.

Our numerical procedure allows to obtain through a realistic approximate solution with real nuclear data. An alternative algorithm, based on Cranck-Nicholson scheme for the equation $h' = \phi$ is proposed in [140].



(a) Neutron flux (red: uncoupled, blue: coupled)



(b) Temperature (in K)

Figure 6.5: Neutron flux and temperature against z (in cm).

Chapter 7

Fluid mechanics

The work presented in this Chapter relates to articles [A9, A5, W1], proceedings [P3, P4] and PhD thesis [210], which references are recalled below:

[A13] *Description and convergence order analysis of the Finite Element-Volume spatial discretization method*, M. A. Puscas, P.-E. Angeli, N. Nouaime, E. Jamelot, International Journal for Numerical Methods in Fluids, 2025..

[A5] *Explicit T-coercivity for the Stokes problem: a coercive finite element discretization*, P. Ciarlet Jr, E. Jamelot, Computers & Mathematics with Applications, **188**, pp. 137–159, 2025.

[A9] *Stability estimates for solving Stokes problem with nonconforming finite elements*, E. Jamelot, Comptes Rendus. Mathématiques, **363**, pp. 115-137, 2025.

[W1] *Convergence analysis of the* $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ *discretization in the TrioCFD code*, P.-E. Angeli, H. Bertrand, E. Jamelot, submitted to Computers & Mathematics With Applications.

[P4] *Improved Crouzeix-Raviart scheme for the Stokes and Navier-Stokes problem*, É. Chénier, E. Jamelot, C. Le Potier, A. Peitavy, New trends in complex flows, 2022, Sept. 19-21, Institut Henri Poincaré, Paris.

[P3] *Improved Crouzeix-Raviart scheme for the Stokes problem*, É. Chénier, E. Jamelot, C. Le Potier, A. Peitavy, Finite Volumes for Complex Applications **10**, 2023, Oct. 30-Nov. 3, Strasbourg.

[210] *Schémas de discrétisation en pression et éléments finis de Crouzeix-Raviart pour les écoulements de fluides incompressibles*, A. Peitavy, PhD thesis, Université Gustave Eiffel, 2024.

This Chapter is organized as follows. In Section 7.1, we recall Navier-Stokes equations and Stokes problem. In Section 7.2, we prove that Stokes problem is well posed using basic T-coercivity. Then, in Section 7.3, we introduce discretization notations and useful properties. In Section 7.4, we study the stability and convergence of Stokes problem discretized with nonconforming finite elements [121, Section 5], say $\mathbf{P}^k_{nc} \times P^{k-1}_{disc}$ for $k \in \mathbb{N}^\star$, using discrete T-coercivity. To do so, we need to exhibit a Fortin operator, that is given for $k = 1$ $d = 2$ or 3 in §7.4.1, and for $k = 2$, $d = 2$ in §7.4.2. We give estimates of the stability constants in both cases. In Section 7.4.3, we illustrate the importance of using a divergence-free velocity on some numerical experiments.

In the TrioCFD code [145, 175, 153], the discrete pressure space is built with P_1-Lagrange plus P_0 basis functions. This leads to the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ mixed finite element method. This method shows interesting numerical results. However, only an incomplete proof of the stability is available in [175, 153]. In Section 7.5, we gather the evidence and to write it as clearly as possible, we derive error estimates and

we prove some super-convergence results obtained in $2D$ [P7, W1]. We also provide numerical results to illustrate its efficiency [A13, W1]. A description of the time discretization scheme of the TrioCFD code is given in Appendix B.1.1.

In Section 7.7, using the T-coercivity, we propose to solve Stokes problem with the $\mathbf{P}^1 \times P_0$ scheme on a stabilized discrete scheme [A5]. We prove convergence and give numerical results showing the interest of the method.

7.1 Context

The TrioCFD code is a computational fluid mechanics (CFD) simulation software developed at the CEA. It is a C++ open source and massively parallel code. It is dedicated to the numerical simulation of turbulent flows for scientific and industrial applications, particularly in the nuclear field [9]. A summary of the numerical scheme and convergence properties is given in Appendix B. To meet the expectations of code users, it is important to understand the discretisation method used in the TrioCFD code, and also to investigate on how to improve it.

The incompressible Navier-Stokes equations are among the most widely used models for describing the motions of a Newtonian liquid or gaseous fluid, with applications in meteorology, aeronautics, the automotive industry and energy production systems. In the case of an isothermal, incompressible fluid[1], these equations can be written in the following dimensionless form:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{u}, p) \text{ such that} \\ \partial_t \mathbf{u} - \nu\, \boldsymbol{\Delta}\, \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla p = \mathbf{f}, \quad \forall (t,\, \mathbf{x}) \in (0,\, T] \times \Omega, \quad \text{(i)} \\ \qquad\qquad \operatorname{div} \mathbf{u} = 0, \quad \forall (t,\, \mathbf{x}) \in (0,\, T] \times \Omega, \quad \text{(ii)} \\ \qquad\qquad \mathbf{u}(\mathbf{x},\, 0) = \mathbf{u}_0, \quad \forall \mathbf{x} \in \Omega. \qquad\qquad \text{(iii)} \end{array}\right. \tag{7.1.1}$$

The domain of study $(0, T) \times \Omega$ is such that $T > 0$ and Ω be domain of $\mathbb{R}^d$, simply connected with a Lipschitz boundary and time independent. We must also define boundary conditions for velocity and boundary conditions or uniqueness condition for pressure. The unknown $\mathbf{u}(\mathbf{x},\, t) = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$ represents the fluid velocity field and the unknown $p(\mathbf{x},\, t) \in \mathbb{R}$ represents the fluid pressure field divided by the fluid density. The constant parameter ν represents the kinematic viscosity. The data $\mathbf{f}(\mathbf{x},\, t) \in \mathbb{R}^d$ represents a source term of body force, divided by the fluid density. Equation (7.1.1)-(i) is called the momentum equation. It corresponds to Newton law, applied to a fluid. Equation (7.1.1)-(ii) corresponds to the conservation of mass equation for incompressible fluids. The term $-\nu\Delta\mathbf{u}$ is called the viscous diffusion term and the nonlinear term $(\mathbf{u}\cdot\nabla)\mathbf{u}$ is called the convection term. It comes from the acceleration term, which is written as $\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \partial_t \mathbf{u} + \mathbf{u}\cdot(\boldsymbol{\nabla}\,\mathbf{u})$. By calculation, we can see that $\mathbf{u}\cdot(\boldsymbol{\nabla}\,\mathbf{u}) = (\mathbf{u}\cdot\nabla)\mathbf{u}$. We refer to [161] and [64] for more details on the study of Navier-Stokes equations.

When the convection term is negligible compared to the viscous diffusion term ($\nu >> 1$), Navier-Stokes equations can be simplified. In the steady-state case, we obtain the following system, called Stokes problem:

$$\text{Find } (\mathbf{u}, p) \text{ such that} \quad -\nu\, \boldsymbol{\Delta}\, \mathbf{u} + \nabla p = \mathbf{f}, \quad \operatorname{div} \mathbf{u} = 0, \tag{7.1.2}$$

endowed with boundary conditions.

The Stokes problem (7.1.2), describes the flow of a stationary incompressible Newtonian fluid. Notice that from the point of view of numerical analysis, the study of this problem represents an essential step in constructing a good discretization of the Navier-Stokes equations.

In TrioCFD code, the spatial discretization of the Navier-Stokes equations is based on the so-called *Crouzeix-Raviart* finite element method [121, Example 4] for which we use the nonconforming linear finite element $\mathbf{P}^1_{nc}$ to discretize the fluid velocity and the P^0 finite element to discretize its pressure. In their seminal paper [121], Crouzeix and Raviart actually exhibited four finite elements, and among

[1] i.e. the fluid density is assumed to be constant

them, two were conforming. Hence, we will write $\mathbf{P}^1_{nc} \times P^0$ scheme instead of Crouzeix-Raviart finite elements. This scheme presents nice feature to solve Stokes problem: it is stable (unlike the $\mathbf{P}^1 \times P^0$ scheme), the velocity mass matrix is diagonal in $2D$, and can be lumped in $3D$. The discrete velocity is divergence free element by element. Moreover, the stencil of the stiffness matrix is equal to 5 in $2D$ and 6 in $3D$, and there is no cross-point, which makes parallel computations easier.Nevertheless, this scheme is often inaccurate when solving Navier-Stokes equations [145, 180]. There exists different cures, such that building a divergence free subspace of the discrete velocity space [174], projecting the test functions on some $\mathbf{H}(\mathrm{div})$-conforming space [194], or adding P^1-Lagrange degrees of freedom for the pressure, as it is done in the TrioCFD code. Another direction consists in writing Stokes problem with three unknowns: the velocity, the pressure, and the vorticity $\omega := \mathbf{curl}\,\mathbf{u}$ [141, 142, 223, 224].

In this Chapter, we explore different directions. In Section 7.4, we use order 2 nonconforming finite elements [A9]. In Section 7.5, we study TrioCFD discretization. In Section 7.6, we consider a finite volume scheme to discretize the pressure [P3, P4]. In Section 7.7, we use stabilized $\mathbf{P}^1 \times P^0$ finite elements as post-treatment [A5].

Let us first prove the well-posedness of the Stokes problem in the light of the T-coercivity theory. We consider here that $\Omega \subset \mathbb{R}^3$, the domain of study, is a Lipschitz polyhedron. We refer to [43] for the study of Stokes problem coupled with Darcy equations in a porous media with cracks.

7.2 Basic T-coercivity for the Stokes problem

For convenience, we consider the generalized Stokes problem (i.e. we may have $\mathrm{div}\,\mathbf{u} \neq 0$):

$$\text{Find } (\mathbf{u}, p) \text{ such that } \; -\nu\,\boldsymbol{\Delta}\,\mathbf{u} + \nabla\, p = \mathbf{f}, \quad \mathrm{div}\,\mathbf{u} = g. \tag{7.2.1}$$

with Dirichlet boundary conditions for the velocity $\mathbf{u}$ and a normalization condition for the pressure p: $p \in L^2_{zmv}(\Omega)$. If the Dirichlet boundary conditions are homogeneous, we write $(7.2.1)_H$, and $(7.2.1)_{NH}$ else. The natural function space for the velocity is $\mathbf{H}^1(\Omega)$ and if homogeneous boundary conditions are prescribed, it is $\mathbf{H}^1_0(\Omega)$, while for the pressure the natural function space is $L^2_{zmv}(\Omega)$. The data is $(\mathbf{f}, g)$. The vector field $\mathbf{f} \in \mathbf{H}^{-1}(\Omega)$ represents a body forces divided by the fluid density. The scalar field $g \in L^2_{zvm}(\Omega)$ is some abstract data that will be useful for the analysis of the case of nonhomogeneous Dirichlet boundary conditions. The variational formulation of Problem $(7.2.1)_H$ reads:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{u}, p) \in \mathbf{H}^1_0(\Omega) \times L^2_{zmv}(\Omega) \text{ such that} \\ \nu(\mathbf{u}, \mathbf{v})_{\mathbf{H}^1_0(\Omega)} - (p, \mathrm{div}\,\mathbf{v})_{L^2(\Omega)} = \langle \mathbf{f}, \mathbf{v}\rangle_{\mathbf{H}^1_0(\Omega)} \quad \forall \mathbf{v} \in \mathbf{H}^1_0(\Omega)\ ; \\ \qquad\qquad (q, \mathrm{div}\,\mathbf{u})_{L^2(\Omega)} = (g, q)_{L^2(\Omega)} \quad \forall q \in L^2_{zmv}(\Omega). \end{array}\right. \tag{7.2.2}$$

Recall $\mathbf{V} := \{\mathbf{v} \in \mathbf{H}^1_0(\Omega) \mid \mathrm{div}\,\mathbf{v} = 0\}$, as defined in Section 4.3. We recall the following regularity result, in the case where $g = 0$, i.e. $\mathbf{u} \in \mathbf{V}$:

Proposition 7.2.1 (`[184], [121, Equation (3.34)]`). *Suppose that Ω is convex. Then, the mapping $(\mathbf{v}, q) \mapsto -\nu\Delta\mathbf{v} + \nabla\, q$ is an isomorphism of $(\mathbf{V} \cap \mathbf{H}^2(\Omega)) \times (L^2_{zmv}(\Omega) \cap H^1(\Omega))$ onto $\mathbf{L}^2(\Omega)$, and it holds:*

$$\nu|\mathbf{v}|_{\mathbf{H}^2(\Omega)} + |p|_{H^1(\Omega)} \leq \|\mathbf{f}\|_{\mathbf{L}^2(\Omega)} \text{ where } \mathbf{f} := -\nu\Delta\mathbf{v} + \nabla\, q.$$

Problem (7.2.2) is a saddle point-problem. Using classical theory, one proves that Problem (7.2.2) is well-posed with the help of Poincaré-Steklov Theorem 4.2.2 and Proposition 4.3.1. Check for instance the proof of [161, Theorem I.5.1]. Let us set $\mathcal{X} = \mathbf{H}^1_0(\Omega) \times L^2_{zmv}(\Omega)$ which is a Hilbert space which we endow with the following norm:

$$\|(\mathbf{v}, q)\|_{\mathcal{X},\nu} = \left(\|\mathbf{v}\|^2_{\mathbf{H}^1_0(\Omega)} + \nu^{-2}\,\|q\|^2_{L^2(\Omega)}\right)^{1/2}. \tag{7.2.3}$$

This norm is chosen to account for physical phenomena due to small viscosity. Typically, if the result of a physics experiment is $(\mathbf{u}, p)$, the ratio of the two components of the norm $\|(\mathbf{u}, p)\|_{\mathcal{X},\nu}$, respectively equal to $\|\mathbf{u}\|_{\mathbf{H}^1_0(\Omega)}$ and $\nu^{-1}\|p\|_{L^2(\Omega)}$, varies linearly with ν.

We define the following bilinear symmetric and continuous form:

$$\left\{\begin{array}{rcl} a : \mathcal{X}\times\mathcal{X} & \to & \mathbb{R} \\ (\mathbf{u}', p')\times(\mathbf{v}, q) & \mapsto & \nu(\mathbf{u}', \mathbf{v})_{\mathbf{H}^1_0(\Omega)} - (p', \operatorname{div}\mathbf{v})_{L^2(\Omega)} - (q, \operatorname{div}\mathbf{u}')_{L^2(\Omega)} \end{array}\right. .$$

We also define the linear and continuous form:

$$\left\{\begin{array}{rcl} \ell : \mathcal{X} & \to & \mathbb{R} \\ (\mathbf{v}, q) & \mapsto & \langle \mathbf{f}, \mathbf{v}\rangle_{\mathbf{H}^1_0(\Omega)} - (g, q)_{L^2(\Omega)} \end{array}\right. .$$

We can write Problem $(7.2.1)_H$ in an equivalent way as follows:

$$\text{Find } (\mathbf{u}, p) \in \mathcal{X} \text{ s.t. } \quad a\left((\mathbf{u}, p), (\mathbf{v}, q)\right) = \ell(\,(\mathbf{v}, q)\,) \quad \forall (\mathbf{v}, q) \in \mathcal{X}. \tag{7.2.4}$$

Let us prove that Problem (7.2.4) is well-posed using basic T-coercivity [26, 60].

Proposition 7.2.2. *The bilinear form $a(\cdot,\cdot)$ is T-coercive.*

Proof. We follow here the proof given in [60, 26]. Let us consider $(\mathbf{u}', p') \in \mathcal{X}$ and let us build $(\mathbf{v}^\star, q^\star) = T(\mathbf{u}', p') \in \mathcal{X}$ satisfying (4.6.2) (with $V = W = \mathcal{X}$). We need three main steps.

1. According to Proposition 4.3.1, there exists $\tilde{\mathbf{v}}_{p'} \in \mathbf{V}^\perp$ such that:

$$\operatorname{div}\tilde{\mathbf{v}}_{p'} = p' \text{ in } \Omega \text{ and } \|\tilde{\mathbf{v}}_{p'}\|_{\mathbf{H}^1_0(\Omega)} \leq C_{\rm div}\,\|p'\|_{L^2(\Omega)}. \tag{7.2.5}$$

   Let us set $(\mathbf{v}^\star, q^\star) := (\lambda\mathbf{u}' - \nu^{-1}\tilde{\mathbf{v}}_{p'}, -\lambda p')$, with $\lambda > 0$ to be fixed. We obtain:

$$a\left((\mathbf{u}', p'), (\mathbf{v}^\star, q^\star)\right) = \nu\,\lambda\|\mathbf{u}'\|^2_{\mathbf{H}^1_0(\Omega)} + \nu^{-1}\,\|p'\|^2_{L^2(\Omega)} - (\mathbf{u}', \tilde{\mathbf{v}}_{p'})_{\mathbf{H}^1_0(\Omega)}. \tag{7.2.6}$$

2. In order to bound the last term of (7.2.6), we use the Young inequality and then inequality (7.2.5), and it follows that

$$\forall \eta > 0, \quad (\mathbf{u}', \tilde{\mathbf{v}}_{p'})_{\mathbf{H}^1_0(\Omega)} \leq \frac{\eta}{2}\|\mathbf{u}'\|^2_{\mathbf{H}^1_0(\Omega)} + \frac{\eta^{-1}}{2}\,(C_{\rm div})^2\,\|p'\|^2_{L^2(\Omega)}. \tag{7.2.7}$$

3. Using the bound (7.2.7) in (7.2.6): $a\left((\mathbf{u}', p'), (\mathbf{v}^\star, q^\star)\right) \geq \left(\nu\,\lambda - \frac{\eta}{2}\right)\|\mathbf{u}'\|^2_{\mathbf{H}^1_0(\Omega)} + \left(\nu^{-1} - \frac{\eta^{-1}}{2}\,(C_{\rm div})^2\right)\|p'\|^2_{L^2(\Omega)}$. We look for $\eta > 0$ such that $2\nu\lambda > \eta$ and $\eta > \frac{\nu}{2}\,(C_{\rm div})^2$, which amounts to requiring $\lambda > \frac{1}{4}\,(C_{\rm div})^2$.

According to the above, provided that $\lambda > \frac{1}{4}\,(C_{\rm div})^2$, we have proved that the operator $T_\lambda \in \mathcal{L}(\mathcal{X})$ defined by $T_\lambda\left((\mathbf{u}', p')\right) = (\lambda\mathbf{u}' - \nu^{-1}\tilde{\mathbf{v}}_{p'}, -\lambda p')$ is such that: $\exists\alpha_\lambda > 0,\ \forall (\mathbf{u}', p') \in \mathcal{X}, a\left((\mathbf{u}', p'), T_\lambda\left((\mathbf{u}', p')\right)\right) \geq \alpha_\lambda\,\|(\mathbf{u}', p')\|^2_{\mathcal{X},\nu}$.
The injectivity of the operator T_λ follows. Given $(\mathbf{v}^\star, q^\star) \in \mathcal{X}$, choosing $(\mathbf{u}', p') = ((\lambda^{-1}\mathbf{v}^\star - \nu^{-1}\lambda^{-2}\tilde{\mathbf{v}}_{q^\star}, -\lambda^{-1}q^\star)$ yields $T_\lambda((\mathbf{u}', p')) = (\mathbf{v}^\star, q^\star)$. Hence, the operator $T_\lambda \in \mathcal{L}(\mathcal{X})$ is bijective. ◇

Remark 7.2.1. *In the above proof, we established that the bijective operator $T_\lambda\left((\mathbf{u}', p')\right) = (\lambda\mathbf{u}' - \nu^{-1}\tilde{\mathbf{v}}_{p'}, -\lambda p')$ leads to T-coercivity as soon as $\lambda > \frac{1}{4}(C_{\rm div})^2$. Observe that one can have even more flexibility in the choice of T by choosing a different factor in front of $\tilde{\mathbf{v}}_{p'}$, and then choosing λ accordingly.*

We can now prove the following result for the generalized Stokes problem:

Theorem 7.2.1 (`[A5, Theorem 2]`). *Problem (7.2.4) is well-posed, so it admits one and only one solution for any $(\mathbf{f}, g) \in \mathbf{H}^{-1}(\Omega)\times L^2_{zmv}(\Omega)$. Writing $\mathbf{u} = \mathbf{u}_0 + \mathbf{u}_\perp$ with $\mathbf{u}_0 \in \mathbf{V}$ and $\mathbf{u}_\perp \in \mathbf{V}^\perp$, the solution is such that:*

$$\forall \mathbf{f} \in \mathbf{H}^{-1}(\Omega),\ \forall g \in L^2_{zmv}(\Omega) \quad \left\{\begin{array}{l} \|\mathbf{u}_\perp\|_{\mathbf{H}^1_0(\Omega)} \leq C_{\rm div}\|g\|_{L^2(\Omega)}, \\ \|\mathbf{u}_0\|_{\mathbf{H}^1_0(\Omega)} \leq \nu^{-1}\,\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)}, \\ \|p\|_{L^2(\Omega)} \leq C_{\rm div}\,\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)} + \nu\,C^2_{\rm div}\,\|g\|_{L^2(\Omega)}. \end{array}\right. \tag{7.2.8}$$

Proof. According to Proposition 7.2.2, the continuous bilinear form $a(\cdot,\cdot)$ is T-coercive. Hence, according to Theorem 4.6.1, Problem (7.2.4) is well-posed. Let us now derive (7.2.8). Consider $(\mathbf{u},p)$ the unique solution of Problem (7.2.4), where $\mathbf{u}=\mathbf{u}_0+\mathbf{u}_\perp$ with $\mathbf{u}_0\in\mathbf{V}$ and $\mathbf{u}_\perp\in\mathbf{V}^\perp$. Choosing $\mathbf{v}=0$, we obtain that $\operatorname{div}\mathbf{u}_\perp=g$, so that $\|\mathbf{u}_\perp\|_{\mathbf{H}^1_0(\Omega)}\leq C_{\rm div}\|g\|_{L^2(\Omega)}$. Now, choosing $\mathbf{v}=\mathbf{u}_0$ and using orthogonality, we have: $\nu\,\|\mathbf{u}_0\|^2_{\mathbf{H}^1_0(\Omega)}=\langle\mathbf{f},\mathbf{u}_0\rangle_{\mathbf{H}^{-1}(\Omega),\mathbf{H}^1_0(\Omega)}\leq\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)}\,\|\mathbf{u}_0\|_{\mathbf{H}^1_0(\Omega)}$, so that: $\|\mathbf{u}_0\|_{\mathbf{H}^1_0(\Omega)}\leq\nu^{-1}\,\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)}$. Next, we choose, in (7.2.4), $q=0$ and $\mathbf{v}=-\tilde{\mathbf{v}}_p\in\mathbf{V}^\perp$, where $\operatorname{div}\tilde{\mathbf{v}}_p=p$ and $\|\tilde{\mathbf{v}}_p\|_{\mathbf{H}^1_0(\Omega)}\leq C_{\rm div}\,\|p\|_{L^2(\Omega)}$ (see Proposition 4.3.1). Since $\mathbf{u}_0\in\mathbf{V}$, it holds that $(\mathbf{u}_0,\tilde{\mathbf{v}}_p)_{\mathbf{H}^1_0(\Omega)}=0$. This gives:

$$\begin{aligned}\|p\|^2_{L^2(\Omega)} &= (p,\operatorname{div}\tilde{\mathbf{v}}_p)_{L^2(\Omega)}=-\langle\mathbf{f},\tilde{\mathbf{v}}_p\rangle_{\mathbf{H}^1_0(\Omega)}+\nu\langle\Delta\mathbf{u}_\perp,\tilde{\mathbf{v}}_p\rangle_{\mathbf{H}^1_0(\Omega)},\\ &\leq \Big(\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)}+\nu\|\mathbf{u}_\perp\|_{\mathbf{H}^1_0(\Omega)}\Big)\,\|\tilde{\mathbf{v}}_p\|_{\mathbf{H}^1_0(\Omega)}\leq C_{\rm div}\,\Big(\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)}+\nu\,C_{\rm div}\,\|g\|_{L^2(\Omega)}\Big)\,\|p\|_{L^2(\Omega)},\end{aligned}$$

so that: $\|p\|_{L^2(\Omega)}\leq C_{\rm div}\,\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)}+\nu\,C^2_{\rm div}\,\|g\|_{L^2(\Omega)}$. ◇

The discretization of Stokes problem poses a well-known difficulty: it is indeed difficult to reproduce the orthogonal splitting $\mathbf{H}^1_0(\Omega)=\mathbf{V}\overset{\perp}{\oplus}\mathbf{V}^\perp$ at a discrete level. Before proposing different discretizations taking into account this difficulty, we introduce some useful notations.

7.3 Discretization notations

We use the notations given in Section 4.7. Consider $(\mathcal{T}_h)_h$ a triangulation sequence of Ω, cf. Section 4.7. For all $f\in\mathcal{I}_F$, M_f denotes the barycentre of F_f, and $\mathbf{n}_f$ its unit normal (outward oriented if $F_f\in\partial\Omega$). We set $\mathcal{S}_f=|F_f|\,\mathbf{n}_f$, and we call h_f the diameter of F_f. For all $\ell\in\mathcal{I}_K$, for all $f\in\mathcal{I}_{F,\ell}$, we call $\mathbf{n}_{f,\ell}$ the normal vector of F_f outgoing of K_ℓ. We call $\hat{K}$ the reference simplex. For all $\ell\in\mathcal{I}_K$, we denote by $T_\ell:\hat{K}\to K_\ell$ the geometric mapping such that $\forall\hat{\mathbf{x}}\in\hat{K}$, $\mathbf{x}_{|K_\ell}=T_\ell(\hat{\mathbf{x}})=\mathbb{B}_\ell\hat{\mathbf{x}}+\vec{OS}_i$, where $\mathbb{B}_\ell\in\mathbb{R}^{d\times d}$ and $i\in\mathcal{I}_{S,\ell}$ so that $\vec{OS}_i\in\mathbb{R}^d$. Let us set $J_\ell=\det(\mathbb{B}_\ell)$. Recall that we have [147, Lemma 11.1]:

$$|J_\ell|=|K_\ell|\,|\hat{K}|^{-1},\quad \|\mathbb{B}_\ell\|=h_\ell\,(\rho_{\hat{K}})^{-1},\quad \|{\mathbb{B}_\ell}^{-1}\|=h_{\hat{K}}\,(\rho_\ell)^{-1}. \tag{7.3.1}$$

For all $v\in L^2(\Omega)$, we set:

$$v_\ell=v_{|K_\ell},\quad \underline{v}_\ell=\int_{K_\ell}v\,/\,|K_\ell|,\quad \hat{v}_\ell=v\circ T_\ell. \tag{7.3.2}$$

By changing the variable, we get:

$$\|v_\ell\|^2_{L^2(K_\ell)}=|J_\ell|\,\|\hat{v}_\ell\|^2_{L^2(\hat{K})}. \tag{7.3.3}$$

By changing the variable, we have:

$$\forall\ell\in\mathcal{I}_K,\ \forall v_\ell\in H^1(K_\ell),\quad \nabla v_\ell=({\mathbb{B}_\ell}^{-1})^T\,\nabla_{\hat{\mathbf{x}}}\,\hat{v}_\ell\ \text{ and }\ \nabla_{\hat{\mathbf{x}}}\,\hat{v}_\ell=(\mathbb{B}_\ell)^T\,\nabla v_\ell. \tag{7.3.4}$$

Hence, for all $\ell\in\mathcal{I}_K$, $\forall v_\ell\in H^1(K_\ell)$, it stands:

$$\|\nabla v_\ell\|^2_{\mathbf{L}^2(K_\ell)} \leq \|{\mathbb{B}_\ell}^{-1}\|^2\,|J_\ell|\,\|\nabla_{\hat{\mathbf{x}}}\,\hat{v}_\ell\|^2_{\mathbf{L}^2(\hat{K})}, \tag{7.3.5}$$

$$\|\nabla_{\hat{\mathbf{x}}}\,\hat{v}_\ell\|^2_{\mathbf{L}^2(\hat{K})} \leq \|\mathbb{B}_\ell\|^2\,|J_\ell|^{-1}\,\|\nabla v_\ell\|^2_{\mathbf{L}^2(K_\ell)}. \tag{7.3.6}$$

Let us recall the Poincaré-Steklov inequality in cells [147, Lemma 12.11]:

$$\forall\ell\in\mathcal{I}_K,\forall v_\ell\in H^1(K_\ell),\quad \|v_\ell-\underline{v}_\ell\|_{L^2(K_\ell)}\leq\pi^{-1}h_\ell\|\nabla v_\ell\|_{\mathbf{L}^2(K_\ell)}. \tag{7.3.7}$$

For all $\ell\in\mathcal{I}_K$ and $f\in\mathcal{I}_{F,\ell}$, we denote by $T_{f,\ell}$ the geometric mapping such that $\forall\mathbf{x}\in F_f$, $T_{f,\ell}(T_\ell^{-1}(\mathbf{x}))=\mathbf{x}$. We let $\hat{F}_{f,\ell}=T_{f,\ell}(F_f)$ and $|J_{f,\ell}|=|F_f|\,|\hat{F}_{f,\ell}|^{-1}$. For all $v\in\mathcal{P}H^1$, for all $\ell\in\mathcal{I}_K$ and $f\in\mathcal{I}_{F,\ell}$, we set:

$$v_{f,\ell}=v_{\ell|F_f},\quad \underline{v}_{f,\ell}=\int_{F_f}v_{f,\ell}\,/\,|F_f|,\quad \hat{v}_{f,\ell}=v\circ T_{f,\ell} \tag{7.3.8}$$

By changing the variable, we get:

$$\|v_{f,\ell}\|^2_{L^2(F_f)} = |F_f|\,|\hat{F}_{f,\ell}|^{-1}\,\|\hat{v}_{f,\ell}\|^2_{L^2(\hat{F}_{f,\ell})}. \tag{7.3.9}$$

Let us set $\mathbf{n}_{\hat{F}_{f,\ell}}$ the unit normal vector of the face $\hat{F}_{f,\ell}$ outgoing from $\hat{K}$. We have :

$$\mathbf{n}_{f,\ell} = |J_\ell|\,|J_{f,\ell}|^{-1}\,(\mathbb{B}_\ell{}^{-1})^T\,\mathbf{n}_{\hat{F}_{f,\ell}}. \tag{7.3.10}$$

We will need [121, Inequality (3.17)]:

$$\forall \ell \in \mathcal{I}_K,\, \forall f \in \mathcal{I}_{F,\ell} \quad |J_{f,\ell}|\,|J_\ell|^{-1} \le \|\mathbb{B}_\ell^{-1}\| \lesssim (\rho_\ell)^{-1}. \tag{7.3.11}$$

Let $k > 0$ and $C_{k,d} = ((k+1)(k+d)/d)^{\frac{1}{2}}$. The following inequality holds [243, Theorem 5]:

$$\forall \ell \in \mathcal{I}_K,\, \forall v_\ell \in P^k(K_\ell), \quad \|v_\ell\|_{L^2(\partial K_\ell)} \le C_{k,d}\,\left(|\partial K_\ell|\,|K_\ell|^{-1}\right)^{\frac{1}{2}}\,\|v_\ell\|_{L^2(K_\ell)}. \tag{7.3.12}$$

Using successively (7.3.12) and (7.3.11), we obtain:

$$\forall \ell \in \mathcal{I}_K, \quad \forall v_\ell \in P^1(K_\ell), \forall f \in \mathcal{I}_{F,\ell}, \quad \|v_\ell - \underline{v}_\ell\|_{L^2(F_f)} \lesssim (\rho_\ell)^{-\frac{1}{2}}\|v_\ell - \underline{v}_\ell\|_{L^2(K_\ell)}. \tag{7.3.13}$$

Proposition 7.3.1. *The following inverse inequality holds:*

$$\forall \ell \in \mathcal{I}_K,\, \forall v_\ell \in P^1(K_\ell), \quad \|\nabla v_\ell\|_{\mathbf{L}^2(K_\ell)} \lesssim (\rho_\ell)^{-1}\|v_\ell - \underline{v}_\ell\|_{L^2(K_\ell)}. \tag{7.3.14}$$

Proof. Let $v_\ell \in P^1(K_\ell)$. Since $\underline{v}_\ell$ is a constant and ∇v is a constant vector, it holds:

$$\|\nabla v_\ell\|^2_{\mathbf{L}^2(K_\ell)} = \|\nabla(v_\ell - \underline{v}_\ell)\|^2_{\mathbf{L}^2(K_\ell)} = |K_\ell|\,|\nabla(v_\ell - \underline{v}_\ell)|^2 = |K_\ell|^{-1}\left|\int_{K_\ell}\nabla(v_\ell - \underline{v}_\ell)\right|^2.$$

We deduce that:

$$\begin{aligned}
\|\nabla v_\ell\|^2_{\mathbf{L}^2(K_\ell)} &= |K_\ell|^{-1}\left|\int_{\partial K_\ell}(v_\ell - \underline{v}_\ell)\mathbf{n}_{|\partial K_\ell}\right|^2 && \text{using an integration by parts,}\\
&\le |K_\ell|^{-1}\left(\int_{\partial K_\ell}|v_\ell - \underline{v}_\ell|\right)^2 && \text{since } \left|\mathbf{n}_{|\partial K_\ell}\right| \le 1.\\
&\le |K_\ell|^{-1}\,|\partial K_\ell|\,\|v_\ell - \underline{v}_\ell\|^2_{L^2(\partial K_\ell)} && \text{using Cauchy-Schwarz,}\\
&\lesssim (\rho_\ell)^{-2}\|v_\ell - \underline{v}_\ell\|^2_{L^2(K_\ell)} && \text{using (7.3.11) and (7.3.13).}
\end{aligned}$$

◇

We need the following estimates inspired by [147, Lemma 12.15]:

Proposition 7.3.2. *For all $\ell \in \mathcal{I}_K$, for all $v \in H^1(K_\ell)$, for all $f \in \mathcal{I}_{F,\ell}$, it holds:*

$$\|v - \underline{v}_\ell\|_{L^2(F_f)} \lesssim (\sigma_\ell\, h_\ell)^{\frac{1}{2}}\,|v|_{H^1(K_\ell)}, \tag{7.3.15}$$

$$\|v - \underline{v}_f\|_{L^2(F_f)} \lesssim (\sigma_\ell\, h_\ell)^{\frac{1}{2}}\,|v|_{H^1(K_\ell)}. \tag{7.3.16}$$

Proof. Let $\pi_0^f \in \mathcal{L}(L^2(F_f),\mathbb{R})$ be the operator such that for all $v \in L^2(F_f)$, $\pi_0^f(v) = \underline{v}_f$. It holds: $\|\pi_0^f\|_{\mathcal{L}(L^2(F_f),\mathbb{R})} \le 1$. Notice that $v - \underline{v}_f = (Id - \pi_0^f)(v - \underline{v}_\ell)$, hence $\|v - \underline{v}_f\|^2_{L^2(F_f)} \le 2\|v_\ell - \underline{v}_\ell\|^2_{L^2(F_f)}$: inequality (7.3.16) is a consequence of (7.3.15). Changing the variable, using the trace theorem in the reference element, and changing the variable again, we obtain that:

$$\|v_\ell - \underline{v}_\ell\|^2_{L^2(F_f)} \lesssim |F_f|\,|K_\ell|^{-1}\,(\|v_\ell - \underline{v}_\ell\|^2_{L^2(K_\ell)} + (h_\ell)^2\,\|\nabla v_\ell\|^2_{\mathbf{L}^2(K_\ell)}).$$

We conclude using the discrete Poincaré-Steklov inequality (7.3.7) and (7.3.11). ◇

We recall a classical approximation estimate: changing the variable, using Bramble-Hilbert/Deny-Lions Lemma [147, Lemma 11.9] in the reference element, changing the variable again, we have that for

$k \geq 0$, for all $m \in \{0, \ldots, k+1\}$, $\ell \in \mathcal{I}_K$, for all $v \in H^{k+1}(K_\ell)$:

$$\inf_{v_h \in P^k(K_\ell)} |v - v_h|_{H^m(K_\ell)} \lesssim \frac{h^{k+1}}{\rho^m} |v|_{H^{k+1}(K_\ell)}. \tag{7.3.17}$$

Recall that π_0 in defined in Section 4.7. By summation over $\ell \in \mathcal{I}_K$, we deduce from (7.3.7) that:

$$\forall v \in \mathcal{P}_h H^1, \quad \|v - \pi_0(v)\|_{L^2(\Omega)} \leq \pi^{-1} h \| \nabla_h v\|_{\mathbf{L}^2(\Omega)}. \tag{7.3.18}$$

7.4 Nonconforming finite elements [A9]

For all $\ell \in \mathcal{I}_K$, for all $i \in \mathcal{I}_{S,\ell}$, $\lambda_{i,\ell}$ denotes the barycentric coordinate related to the vertex S_i in K_ℓ. We can endow $\mathcal{P}^1(\mathcal{T}_h)$ with the basis $(\phi_i)_{i \in \mathcal{I}_S}$ such that:

$$\forall \ell \in \mathcal{I}_K, \, \phi_{i|K_\ell} = \left\{ \begin{array}{cc} \lambda_{i,\ell} & \text{if } i \in \mathcal{I}_{S,\ell} \\ 0 & \text{otherwise} \end{array} \right. , \text{ so that } \mathcal{P}^1(\mathcal{T}_h) = \text{vect}\left((\phi_i)_{i \in \mathcal{I}_S}\right).$$

The nonconforming finite element method was introduced by Crouzeix and Raviart in [121] to solve Stokes problem (7.2.1). We approximate the vector space $\mathbf{H}^1(\Omega)$ component by component by piecewise polynomials of order $k > 0$.

Let us consider X_h (resp. $X_{0,h}$), the space of nonconforming approximation of $H^1(\Omega)$ (resp. $H^1_0(\Omega)$) of order k:

$$\begin{aligned} X_h &= \left\{ v_h \in P^k_{disc}(\mathcal{T}_h); \quad \forall f \in \mathcal{I}^i_F, \forall q_h \in P^{k-1}(F_f), \int_{F_f} [v_h]_{F_f} \, q_h = 0 \right\}; \\ X_{0,h} &= \left\{ v_h \in X_h; \quad \forall f \in \mathcal{I}^b_F, \forall q_h \in P^{k-1}(F_f), \int_{F_f} v_h \, q_h = 0 \right\}. \end{aligned} \tag{7.4.1}$$

The condition on the jumps of v_h on the inner facets is often called the patch-test condition. Due to the patch test, one can prove the following results, which are useful for the numerical analysis of nonconforming finite elements. In the spirit of [130, §1.3], we let $X^\star_{0,h} = X_{0,h} + H^1_0(\Omega)$. It holds:

Proposition 7.4.1. *Let $(v, q) \in X^\star_{0,h} \times H^1(\Omega)$. We have:*

$$\left| \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (v_\ell \, , \, q)_{L^2(F_f)} \right| \lesssim \sigma \, h \, \|v\|_h \, \| \nabla \, q\|_{\mathbf{L}^2(\Omega)}. \tag{7.4.2}$$

Let $(v, q) \in X^\star_{0,h} \times \mathcal{P}^1(\mathcal{T}_h)$. We have:

$$\left| \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (v_\ell \, , \, q)_{L^2(F_f)} \right| \lesssim \sigma \, \|v\|_h \, \|q - \pi_0(q)\|_{L^2(\Omega)}. \tag{7.4.3}$$

Proof. We are inspired by the proof of [42, Lemma 3.1]. Let $(v, q) \in X_{0,h} \times H^1(\Omega)$. It holds:

$$\sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\underline{v}_f, q)_{L^2(F_f)} = \sum_{f \in \mathcal{I}^i_F} \int_{L^2(F_f)} [\underline{v}_f \, q]_f = 0 \text{ and } \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (v_\ell - \underline{v}_f, \underline{q}_\ell)_{L^2(F_f)} = 0.$$

By summation and using Cauchy-Schwarz, we deduce that:

$$\sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (v_\ell, q)_{L^2(F_f)} = \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (v_\ell - \underline{v}_f, q - \underline{q}_\ell)_{L^2(F_f)} \leq \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} \|v_\ell - \underline{v}_f\|_{L^2(F_f)} \, \|q - \underline{q}_\ell\|_{L^2(F_f)}.$$

Using (7.3.15) and (7.3.16), we have: $\|v_\ell - \underline{v}_f\|_{L^2(F_f)} \|q - \underline{q}_\ell\|_{L^2(F_f)} \leq \sigma_\ell\, h_\ell \,\|\nabla v_h\|_{\mathbf{L}^2(K_\ell)} \,\|\nabla q\|_{\mathbf{L}^2(K_\ell)}$. For $q \in \mathcal{P}^1(\mathcal{T}_h)$, using (7.3.13), we get: $\|v_\ell - \underline{v}_f\|_{L^2(F_f)} \|q - \underline{q}_\ell\|_{L^2(F_f)} \leq \sigma_\ell \,\|\nabla v_h\|_{\mathbf{L}^2(K_\ell)} \,\|q - \pi_0(q)\|_{L^2(K_\ell)}$. We conclude using the discrete Cauchy-Schwarz inequality. ⋄

The space $X_{0,h}$ is not a subspace of $H^1_0(\Omega)$, but we can prove a discrete Poincaré-Steklov inequality, following the proof of [226, Theorem D.1]. The proof of [148, Lemma 36.6] is similar, but the vector $\mathbf{s}$ defined in Corollary 4.3.1 is constructed in [148, Lemma 36.6] as the gradient of a scalar function, so that it gives a lower estimate when Ω is non-convex. Alternative proofs are given in [68, 242].

Proposition 7.4.2. *The following discrete Poincaré–Steklov inequality holds:*

$$\forall v_h \in X_{0,h}, \quad \|v_h\|_{L^2(\Omega)} \lesssim \sigma\, C_\Omega\, h_\Omega\, \|v_h\|_h. \tag{7.4.4}$$

Proof. Let $v_h \in X_{0,h}$. According to Corollary 4.3.1, it exists $\mathbf{s} \in \mathbf{H}^1(\Omega)$ such that:

$$\operatorname{div} \mathbf{s} = v_h \text{ and } \|\mathbf{s}\|_{\mathbf{L}^2(\Omega)} + h_\Omega \|\boldsymbol{\nabla}\, \mathbf{s}\|_{\mathbb{L}^2(\Omega)} \leq C_\Omega\, h_\Omega \|v_h\|_{L^2(\Omega)} \tag{7.4.5}$$

We have, by integration by parts:

$$\|v_h\|^2_{L^2(\Omega)} = (v_h, \operatorname{div} \mathbf{s})_{L^2(\Omega)} = -\sum_{\ell \in \mathcal{I}_K} (\nabla\, v_h, \mathbf{s})_{\mathbf{L}^2(K_\ell)} + \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (v_h, \mathbf{s} \cdot \mathbf{n}_{f,\ell})_{L^2(F_f)}. \tag{7.4.6}$$

The first term can be bounded as follows: $(\nabla\, v_h, \mathbf{s})_{\mathbf{L}^2(K_\ell)} \leq \|\nabla\, v_h\|_{\mathbf{L}^2(K_\ell)} \,\|\mathbf{s}\|_{\mathbf{L}^2(K_\ell)}$. Using (7.4.2), we have:

$$\sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (v_h, \mathbf{s} \cdot \mathbf{n}_{f,\ell})_{L^2(F_f)} \leq \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} \sigma_\ell\, h_\ell \,\|\nabla\, v_h\|_{\mathbf{L}^2(K_\ell)} \,\|\boldsymbol{\nabla}\, \mathbf{s}\|_{\mathbb{L}^2(K_\ell)}.$$

Combining these two last results, (7.4.6) now reads: $\|v_h\|^2_{L^2(\Omega)} \lesssim \sum_{\ell \in \mathcal{I}_K} \|\nabla\, v_h\|_{\mathbf{L}^2(K_\ell)} \left(\|\mathbf{s}\|_{\mathbf{L}^2(K_\ell)} + \sigma_\ell\, h_\ell \,\|\boldsymbol{\nabla}\, \mathbf{s}\|_{\mathbb{L}^2(K_\ell)}\right)$.
We obtain (7.4.4) using the discrete Cauchy-Schwarz inequality and (7.4.5). ⋄

As a consequence of Proposition 7.4.2, we have the

Proposition 7.4.3. *The broken norm $v_h \to \|v_h\|_h$ is a norm over $X_{0,h}$.*

Let $\mathcal{J}_h : X_{0,h} \to Y_{0,h}$, with $Y_{0,h} = \mathbf{H}^1_0(\Omega) \cap \mathcal{P}^k_{disc}(\mathcal{T}_h)$ be the averaging operator defined in [147, §22.4.1]. There exists a constant $C^{nc}_{\mathcal{J}_h} \approx \sigma$ independent of h such that:

$$\forall v_h \in X_{0,h}, \quad \|\mathcal{J}_h v_h\|_{H^1_0(\Omega)} \lesssim C^{nc}_{\mathcal{J}_h} \|v_h\|_h. \tag{7.4.7}$$

The space of nonconforming approximation of $\mathbf{H}^1(\Omega)$ (resp. $\mathbf{H}^1_0(\Omega)$) of order k is $\mathbf{X}_h = (X_h)^d$ (resp. $\mathbf{X}_{0,h} = (X_{0,h})^d$). We set $\mathcal{X}_h := \mathbf{X}_{0,h} \times Q_h$ where $Q_h = P^{k-1}_{disc}(\mathcal{T}_h) \cap L^2_{zmv}(\Omega)$. which is a Hilbert space which we endow with the following norm: $\|(\mathbf{v}_h, q_h)\|_{\mathcal{X}_h,\nu} = \left(\|\mathbf{v}_h\|^2_h + \nu^{-2} \|q_h\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}}$. We consider the discrete continuous bilinear form $a_h(\cdot,\cdot)$ such that:

$$\left\{ \begin{array}{rcl} a_h : \mathcal{X}_h \times \mathcal{X}_h & \to & \mathbb{R} \\ (\mathbf{u}'_h, p'_h) \times (\mathbf{v}_h, q_h) & \mapsto & \nu(\mathbf{u}'_h, \mathbf{v}_h)_h - (\operatorname{div}_h \mathbf{v}_h, p'_h)_{L^2(\Omega)} - (\operatorname{div}_h \mathbf{u}'_h, q_h)_{L^2(\Omega)} \end{array} \right. .$$

Let us set $\mathbf{V}_{0,h}$ the discrete space of discrete divergence-free velocities:

$$\mathbf{V}_{0,h} := \left\{ \mathbf{v}_h \in \mathbf{X}_{0,h} \,|\, \forall q_h \in Q_h,\, (\operatorname{div}_h \mathbf{v}_h, q_h)_{L^2(\Omega)} = 0 \right\}. \tag{7.4.8}$$

We recall that [79, Lemma 3.1]:

Proposition 7.4.4. *For all $\mathbf{v}_h \in \mathbf{V}_{0,h}$, for all $\ell \in \mathcal{I}_K$, $\operatorname{div} \mathbf{v}_{h|K_\ell} = 0$.*

Let $\ell_{\mathbf{f}} \in \mathcal{L}(\mathcal{X}_h, \mathbb{R})$ be such that for all $(\mathbf{v}_h, q_h) \in \mathcal{X}_h$:

$$\begin{array}{ll} \text{If } \mathbf{f} \in \mathbf{L}^2(\Omega): & \ell_{\mathbf{f}}\ (\,(\mathbf{v}_h, q_h)\,) = (\mathbf{f}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}. \\ \text{If } \mathbf{f} \in \mathbf{H}^{-1}(\Omega): & \ell_{\mathbf{f}}\ (\,(\mathbf{v}_h, q_h)\,) = \langle \mathbf{f}, \mathcal{J}_h(\mathbf{v}_h)\rangle_{\mathbf{H}^1_0(\Omega)}. \end{array} \tag{7.4.9}$$

The nonconforming discretization of Problem (7.2.4) reads:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{u}_h, p_h) \in \mathcal{X}_h \text{ such that} \\ \quad a_h\ ((\mathbf{u}_h, p_h), (\mathbf{v}_h, q_h)) = \ell_{\mathbf{f}}\ (\,(\mathbf{v}_h, q_h)\,) \quad \forall (\mathbf{v}_h, q_h) \in \mathcal{X}_h. \end{array} \right. \tag{7.4.10}$$

To prove that Problem (7.4.10) is well-posed, we use the T-coercivity theory, with the help of a Fortin operator. This operator will be explained latter, using the discrete basis functions. The discrete stability constant depends on this operator (hence polynomial order k).

Proposition 7.4.5 (`Discrete T-coercivity`). *Suppose that there exists a Fortin operator $\Pi_{nc} \in \mathcal{L}(\mathbf{H}^1(\Omega), \mathbf{X}_h)$ such that*

$$\exists C_{nc}\,|\,\forall \mathbf{v} \in \mathbf{H}^1(\Omega) \qquad \|\Pi_{nc}\mathbf{v}\|_h \leq C_{nc}\|\boldsymbol{\nabla}\,\mathbf{v}\|_{\mathbb{L}^2(\Omega)}, \tag{7.4.11}$$

$$\forall \mathbf{v} \in \mathbf{H}^1(\Omega) \quad (\mathrm{div}_h\,\Pi_{nc}\mathbf{v}, q_h)_{L^2(\Omega)} = (\mathrm{div}\,\mathbf{v}, q_h)_{L^2(\Omega)}, \quad \forall q \in Q_h, \tag{7.4.12}$$

where the constant C_{nc} does not depend on h. Then, the bilinear form $a_h(\cdot,\cdot)$ is T-coercive.

Remark 7.4.1. *We will exhibit C_{nc}, obtaining $C_{nc} = 1$ for $k = 1$, $d = 2, 3$ in §7.4.1, and $C_{nc} = \sigma^2 + 1$ for $k = 2$, $d = 2$ in §7.4.2 .*

Proof. We will follow the proof of Proposition 7.2.2. Let us consider $(\mathbf{u}'_h, p'_h) \in \mathcal{X}_h$ and let us build $(\mathbf{v}^\star_h, q^\star_h) = T_h(\mathbf{u}'_h, p'_h) \in \mathcal{X}_h$ satisfying (4.6.2) (with $V = W = \mathcal{X}_h$). We need three main steps.

1. According to Proposition 4.3.1, there exists $\tilde{\mathbf{v}}_{p'_h} \in \mathbf{V}^\perp$ such that: $\mathrm{div}\,\tilde{\mathbf{v}}_{p'_h} = p'_h$ and $\|\tilde{\mathbf{v}}_{p'_h}\|_{\mathbf{H}^1_0(\Omega)} \leq C_{\mathrm{div}}\,\|p'_h\|_{L^2(\Omega)}$. Consider $\mathbf{v}_{h,p'_h} = \Pi_{nc}\mathbf{v}_{p'_h}$, for all $q_h \in Q_h$, we have: $(\mathrm{div}_h\,\mathbf{v}_{h,p'_h}, q_h)_{L^2(\Omega)} = \nu^{-1}\,(p'_h, q_h)_{L^2(\Omega)}$ and

$$\|\mathbf{v}_{h,p'_h}\|_h \leq \nu^{-1}\,C^{nc}_{\mathrm{div}}\,\|p'_h\|_{L^2(\Omega)} \text{ where } C^{nc}_{\mathrm{div}} = C_{\mathrm{div}}\,C_{nc}. \tag{7.4.13}$$

   Let us set $(\mathbf{v}^\star_h, q^\star_h) := (\lambda_{nc}\mathbf{u}'_h - \nu^{-1}\tilde{\mathbf{v}}_{h,p'}, -\lambda_{nc}p'_h)$, with $\lambda_{nc} > 0$ to be fixed. We obtain:

$$a_h\,\big((\mathbf{u}'_h, p'_h), (\mathbf{v}^\star_h, q^\star_h)\big) = \nu\,\lambda_{nc}\|\mathbf{u}'_h\|^2_h + \nu^{-1}\,\|p'_h\|^2_{L^2(\Omega)} - (\mathbf{u}'_h, \tilde{\mathbf{v}}_{h,p'_h})_h. \tag{7.4.14}$$

2. In order to bound the last term of (7.4.14), we use the Young inequality and then inequality (7.4.13), and it follows that for all $\eta_{nc} > 0$:

$$(\mathbf{u}'_h, \tilde{\mathbf{v}}_{h,p'})_h \leq \frac{\eta_{nc}}{2}\|\mathbf{u}'_h\|^2_h + \frac{\eta^{-1}_{nc}}{2}\,(C^{nc}_{\mathrm{div}})^2\,\|p'_h\|^2_{L^2(\Omega)}. \tag{7.4.15}$$

3. Using the bound (7.4.15) in (7.4.14): $a_h\,\big((\mathbf{u}'_h, p'_h), (\mathbf{v}^\star_h, q^\star_h)\big) \geq \left(\nu\,\lambda_{nc} - \frac{\eta_{nc}}{2}\right)\|\mathbf{u}'_h\|^2_h + \left(\nu^{-1} - \frac{\eta^{-1}_{nc}}{2}\,(C^{nc}_{\mathrm{div}})^2\right)\|p'_h\|^2_{L^2(\Omega)}$.
   We look for $\eta_{nc} > 0$ such that $2\nu\lambda_{nc} > \eta_{nc}$ and $\eta_{nc} > \frac{\nu}{2}\,(C^{nc}_{\mathrm{div}})^2$, which amounts to requiring $\lambda_{nc} > \frac{1}{4}\,(C^{nc}_{\mathrm{div}})^2$.

According to the above, provided that $\lambda_{nc} > \frac{1}{4}\,(C^{nc}_{\mathrm{div}})^2$, we have proved that the operator $T_h \in \mathcal{L}(\mathcal{X}_h)$ defined by $T_h\,\big((\mathbf{u}'_h, p'_h)\big) = (\lambda_{nc}\mathbf{u}'_h - \nu^{-1}\tilde{\mathbf{v}}_{h,p'_h}, -\lambda_{nc}p'_h)$ is such that: $\exists \alpha_{\lambda_{nc}} > 0,\ \forall (\mathbf{u}'_h, p'_h) \in \mathcal{X}_h, a_h\,\big((\mathbf{u}'_h, p'_h), T_h\,\big((\mathbf{u}'_h, p'_h)\big)\big) \geq \alpha_{\lambda_{nc}}\,\|(\mathbf{u}'_h, p'_h)\|^2_{\mathcal{X}_h,\nu}$.
The injectivity of the operator $T_{h,\lambda_{nc}}$ follows. Given $(\mathbf{v}^\star_h, q^\star_h) \in \mathcal{X}_h$, choosing $(\mathbf{u}'_h, p'_h) = ((\lambda^{-1}_{nc}\mathbf{v}^\star_h - \nu^{-1}\lambda^{-2}_{nc}\tilde{\mathbf{v}}_{h,q^\star_h}, -\lambda^{-1}_{nc}q^\star_h)$ yields $T_h((\mathbf{u}'_h, p'_h)) = (\mathbf{v}^\star_h, q^\star_h)$. Hence, the operator $T_h \in \mathcal{L}(\mathcal{X}_h)$ is bijective. ⋄

We can now prove the discrete counterpart of Theorem 7.2.1.

Theorem 7.4.1 (`Discrete well-posedness`). *Suppose that there exists a Fortin operator Π_{nc} satifying* (7.4.11)-(7.4.12). *Then the bilinear form $a_h(\cdot,\cdot)$ is T-coercive. Problem* (7.4.10) *is well-posed. It admits one and only one solution such that:*

$$\begin{array}{ll} \text{if } \mathbf{f} \in \mathbf{L}^2(\Omega): & \|\mathbf{u}_h\|_h \lesssim \nu^{-1}\,C^{nc}_0\,\|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\,, \quad \|p_h\|_{L^2(\Omega)} \lesssim 2\,C^{nc}_0\,C^{nc}_{\mathrm{div}}\,\|\mathbf{f}\|_{\mathbf{L}^2(\Omega)} \\ \text{if } \mathbf{f} \in \mathbf{H}^{-1}(\Omega): & \|\mathbf{u}_h\|_h \lesssim \nu^{-1}\,C^{nc}_{\mathcal{J}_h}\,\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)}, \quad \|p_h\|_{L^2(\Omega)} \lesssim 2\,C^{nc}_{\mathcal{J}_h}\,C^{nc}_{\mathrm{div}}\,\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)} \end{array}, \tag{7.4.16}$$

where $C_0^{nc} = \sigma\, C_\Omega\, h_\Omega$.

Proof. Consider $(\mathbf{u}_h, p_h)$ the unique solution of Problem (7.4.10). Choosing $\mathbf{v}_h = 0$, we obtain that $\operatorname{div}_h \mathbf{u}_h = 0$. Let $\mathbf{f} \in \mathbf{L}^2(\Omega)$. Now, choosing $\mathbf{v}_h = \mathbf{u}_h$ in (7.4.10), using Cauchy-Schwarz inequality, we get that: $\|\mathbf{u}_h\|_h \leq \nu^{-1}\, \sigma\, C_\Omega\, h_\Omega\, \|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}$ using inequality (7.4.4). Consider $(\mathbf{v}_h, q_h) = (\mathbf{v}_{h,p_h}, 0)$ in (7.4.10), where $\mathbf{v}_{h,p_h} = \Pi_{nc}\mathbf{v}_{p_h}$ is built as $\mathbf{v}_{h,p'_h}$ in point 1, setting $p'_h = p_h$. Suppose that $\mathbf{f} \in \mathbf{L}^2(\Omega)$. Notice that $\nu^{-1}\, \|p_h\|^2_{L^2(\Omega)} = \nu\, (\mathbf{u}_h, \mathbf{v}_{h,p_h})_h - (\mathbf{f}, \mathbf{v}_{h,p_h})_{\mathbf{L}^2(\Omega)}$. Using Cauchy-Schwarz inequality, we have: $\nu^{-1}\, \|p_h\|^2_{L^2(\Omega)} \leq \nu\, \|\mathbf{u}_h\|_h\, \|\mathbf{v}_{h,p_h}\|_h + \|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\, \|\mathbf{v}_{h,p_h}\|_{\mathbf{L}^2(\Omega)}$. Using Poincaré-Steklov inequality (7.4.4), Fortin stability (7.4.11), and the previous estimate on $\|\mathbf{u}_h\|_h$, we have:

$$\|p_h\|^2_{L^2(\Omega)} \lesssim 2\, \sigma\, C_\Omega\, h_\Omega\, \|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\, \|\mathbf{v}_{h,p_h}\|_h \lesssim 2\, \sigma\, C_\Omega\, h_\Omega\, C^{nc}_{\mathrm{div}}\, \|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\, \|p_h\|_{L^2(\Omega)}.$$

Let $\mathbf{f} \in \mathbf{H}^{-1}(\Omega)$. We apply the same reasoning, using inequalities (7.4.7) and (7.4.4). ◇

As a corollary of Theorem 7.4.1, we have the following a priori errors estimates:

Theorem 7.4.2 (`[121, Theorem 3, Theorem 4, Theorem 6] and [152, Equation (47)]`). *Under the assumption of Theorem 7.4.1, suppose that* $(\mathbf{u}, p) \in (\mathbf{H}^{1+k}(\Omega) \cap \mathbf{H}^1_0(\Omega)) \times (H^k(\Omega) \cap L^2_{zmv}(\Omega))$*, we have the estimates:*

$$\|\mathbf{u} - \mathbf{u}_h\|_h \lesssim \sigma^l\, h^k\, \left(|\mathbf{u}|_{\mathbf{H}^{k+1}(\Omega)} + \nu^{-1}\, |p|_{H^k(\Omega)}\right), \tag{7.4.17}$$

$$\nu^{-1}\, \|p - p_h\|_{L^2(\Omega)} \lesssim \sigma^l\, h^k\, \left(|\mathbf{u}|_{\mathbf{H}^{k+1}(\Omega)} + \nu^{-1}\, |p|_{H^k(\Omega)}\right). \tag{7.4.18}$$

Suppose moreover that the domain Ω *is convex. Then we have:*

$$\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma^{2\,l}\, h^{k+1}\, \left(|\mathbf{u}|_{\mathbf{H}^{k+1}(\Omega)} + \nu^{-1}\, |p|_{H^k(\Omega)}\right), \tag{7.4.19}$$

The hidden constants depend on k *but they don't depend on mesh. The parameter* σ *is the shape regularity parameter and the exponent* $l > 0$ *depends on* k*. When* $k = 1$*,* $d = 2$*,* 3*, we have:* $l = 1$ *and when* $k = 2$*,* $d = 2$*, we have:* $l = 2$*.*

Remark 7.4.2. *When* Ω *is nonconvex, if the solution to Problem* (7.2.1) *is such that* $(\mathbf{u}, p) \in \mathbf{H}^2(\Omega) \times H^1(\Omega)$*, one can prove that* $\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma^2\, h^{1+s}\, \|(\mathbf{u}, p)\|_\nu$ *for all* $s \in]\frac{1}{2}, s_\Omega[$*, where* $s_\Omega \in]\frac{1}{2}, 1[$*, cf. [124, 1.A].*

The main issue with nonconforming mixed finite elements is the construction of the basis functions. In a recent paper, Sauter explains such a construction in two dimensions [227, Theorem 1.3], and gives a bound to the discrete counterpart $\beta_{\mathcal{T}}(\Omega)$ of $\beta(\Omega)$ defined in (4.3.3):

$$\beta_{\mathcal{T}}(\Omega) = \inf_{q_h \in Q_h \setminus \{0\}} \sup_{\mathbf{v}_h \in \mathbf{X}_{0,h} \setminus \{0\}} \frac{(\operatorname{div}_h \mathbf{v}_h, q_h)_{L^2(\Omega)}}{\|q_h\|_{L^2(\Omega)}\, \|\mathbf{v}_h\|_h} \geq c_{\mathcal{T}}\, (\log(k+1))^{-\alpha}, \tag{7.4.20}$$

where the parameter α is explicit and depends on k and on the mesh topology; and the constant $c_{\mathcal{T}}$ depends only on the shape-regularity of the mesh.

We will now look at different nonconforming discretizations, and provide some numerical results. Hence, we define below our errors estimators.

Definition 7.4.1. *Let* $\mathcal{I}_{h,nc}$ *be the interpolation operator from* $\mathcal{C}^0(\overline{\Omega})$ *to* X_h*. Let* $\mathcal{I}_{h,dg}$ *be the projection operator from* $\mathcal{C}^0(\overline{\Omega})$ *to* $\mathcal{P}^k_{disc}(\mathcal{T}_h)$*,* $k \geq 0$*. Let* $\delta_h \mathbf{u} = \mathcal{I}_{h,nc}\mathbf{u} - \mathbf{u}_h$ *and* $\delta_h p = \mathcal{I}_{h,dg} p - p_h$*, where* $(\mathbf{u}, p)$ *solves* (7.2.1) *and* $(\mathbf{u}_h, p_h) \in \mathbf{X}_h \times \mathcal{P}^k_{disc}(\mathcal{T}_h)$ *is a discrete approximation of* $(\mathbf{u}, p)$*. In our numerical results, we will give the velocity error* $\varepsilon^\nu_0(\mathbf{u}_h)$ *in* $\mathbf{L}^2(\Omega)$*-norm and the pressure error* $\varepsilon^\nu_0(p_h)$ *in* $L^2(\Omega)$*-norm:*

$$\varepsilon^\nu_0(\mathbf{u}_h) = \|\delta_h \mathbf{u}\|_{\mathbf{L}^2(\Omega)} / \|(\mathbf{u}, p)\|_{\mathcal{X},\nu}, \quad \varepsilon^\nu_0(p_h) = \nu^{-1}\, \|\delta_h p\|_{L^2(\Omega)} / \|(\mathbf{u}, p)\|_{\mathcal{X},\nu}. \tag{7.4.21}$$

We also define the velocity error $\varepsilon^\nu_1(\mathbf{u}_h)$ *in* $\mathbf{X}_h$*-norm*

$$\varepsilon^\nu_1(\mathbf{u}_h) = \|\delta_h \mathbf{u}\|_h / \|(\mathbf{u}, p)\|_{\mathcal{X},\nu}. \tag{7.4.22}$$

7.4.1 The $\mathbf{P}^1_{nc} \times P^0$ finite element

We study the linear nonconforming Crouzeix-Raviart mixed finite elements [121, Example 4], usually simply called *Crouzeix-Raviart finite elements* even though Crouzeix and Raviart analysed four types of mixed finite elements in their seminal paper [121]. Let us consider $\mathbf{X}_{cr} = (X_{cr})^d$ (resp. $\mathbf{X}_{0,cr} = (X_{0,cr})^d$), the space of linear nonconforming approximation of $\mathbf{H}^1(\Omega)$ (resp. $\mathbf{H}^1_0(\Omega)$) where:

$$\begin{aligned} X_{cr} &= \left\{ v_h \in P^1_{disc}(\mathcal{T}_h)\,; \quad \forall f \in \mathcal{I}^i_F, \int_{F_f} [v_h]_{F_f} = 0 \right\}; \\ X_{0,cr} &= \left\{ v_h \in X_{cr}\,; \quad \forall f \in \mathcal{I}^b_F, \int_{F_f} v_h = 0 \right\}. \end{aligned} \tag{7.4.23}$$

For all $\ell \in \mathcal{I}_K$, for all $f \in \mathcal{I}_{F,\ell}$, $\lambda_{f,\ell}$ denotes the barycentric coordinate related to the vertex $S_{f,\ell}$ opposite to face F_f in K_ℓ. We endow X_{cr} with the basis $(\psi_f)_{f\in\mathcal{I}_F}$ such that:

$$\forall \ell \in \mathcal{I}_K,\ \psi_{f|K_\ell} = \begin{cases} 1 - d\lambda_{f,\ell} & \text{if } f \in \mathcal{I}_{F,\ell} \\ 0 & \text{otherwise} \end{cases}, \text{ so that } X_{cr} = \operatorname{vect}\left((\psi_f)_{f\in\mathcal{I}_F}\right).$$

Note that: $\int_{F_{f'}} \psi_f = |F_f|\delta_{f,f'}$, $[\psi_f]_f = 0$ if $f \in \mathcal{I}^i_F$ and $X_{0,cr} = \operatorname{vect}\left((\psi_f)_{f\in\mathcal{I}^i_F}\right)$.

For all $\ell \in \mathcal{I}_K$, we define $\pi^\ell_{cr} : H^1(K_\ell) \to P^1(K_\ell)$ such that:

$$\forall \ell \in \mathcal{I}_K,\ \forall v \in H^1(K_\ell), \quad \pi^\ell_{cr} v = \sum_{f\in\mathcal{I}_{F,\ell}} \underline{v}_f \psi_{f|K_\ell}. \tag{7.4.24}$$

Notice that for all $\ell \in \mathcal{I}_K$, for all $v \in P^1(K_\ell)$, $\pi^\ell_{cr} v = v$. We recall the following result.

Proposition 7.4.6 ([13, Lemma 2])**.** *The norm of the operator π^ℓ_{cr} is independent of K_ℓ:*

$$\forall \ell \in \mathcal{I}_K,\ \forall v \in H^1(K_\ell), \quad \|\nabla \pi^\ell_{cr} v\|_{\mathbf{L}^2(K_\ell)} \le \|\nabla v\|_{\mathbf{L}^2(K_\ell)}. \tag{7.4.25}$$

Hence: $\|\pi^\ell_{cr}\|_{\mathcal{L}(H^1(K_\ell),P^1(K_\ell))} \le 1$.

Proof. Using integration by parts, it holds: $|K_\ell|\,\nabla \pi^\ell_{cr} v = \int_{K_\ell} \nabla \pi^\ell_{cr} v = \int_{\partial K_\ell} \pi^\ell_{cr} v \mathbf{n}_{|\partial K_\ell} = \int_{\partial K_\ell} v \mathbf{n}_{|\partial K_\ell} = \int_{K_\ell} \nabla v$. ⋄

In order to prove convergence estimates, we need the following result.

Proposition 7.4.7. *The following estimates hold, for all $\ell \in \mathcal{I}_K$, for all $s \in [0,1]$, for all $v \in H^{1+s}(K_\ell)$:*

$$\|v - \pi^\ell_{cr} v\|_{L^2(K_\ell)} \le 2\,(h_\ell)^{1+s}\,|v|_{H^{1+s}(K_\ell)}, \tag{7.4.26}$$

$$\|\nabla(v - \pi^\ell_{cr} v)\|_{\mathbf{L}^2(K_\ell)} \le 2\,\sigma_\ell\,(h_\ell)^{s}\,|v|_{H^{1+s}(K_\ell)}, \tag{7.4.27}$$

$$\|v - \pi^\ell_{cr} v\|_{L^2(\partial K_\ell)} \le 2\,(\sigma_\ell\, h_\ell)^{\frac{1}{2}}\,(h_\ell)^{s}\,|v|_{H^{1+s}(K_\ell)}. \tag{7.4.28}$$

Proof. Since for all $v_h \in P^1(K_\ell)$, $v - \pi^\ell_{cr} v = (Id - \pi^\ell_{cr})(v - v_h)$, we have:

$$\|v - \pi^\ell_{cr} v\|_{H^1(K_\ell)} \le \|Id - \pi^\ell_{cr}\|_{\mathcal{L}(H^1(K_\ell),P^1(K_\ell))} \inf_{v_h\in P^1(K_\ell)} \|v - v_h\|_{H^1(K_\ell)} \le 2 \inf_{v_h\in P^1(K_\ell)} \|v - v_h\|_{H^1(K_\ell)}.$$

Using estimate (7.3.17), we deduce that, for $m = 0, 1$:

$$\|v - \pi^\ell_{cr} v\|_{L^2(K_\ell)} \le 2\,(h_\ell)^{1+m}\,|v|_{H^{1+m}(K_\ell)} \quad \text{and} \quad \|\nabla(v - \pi^\ell_{cr} v)\|_{\mathbf{L}^2(K_\ell)} \le 2\,\sigma_\ell (h_\ell)^m\,|v|_{H^{1+m}(K_\ell)}.$$

We obtain (7.4.26) and (7.4.27), using interpolation property [237, Lemma 22.2] We estimate $\|v - \pi^\ell_{cr} v\|_{L^2(\partial K_\ell)}$ with the same tools, changing the variable, using the trace theorem in the reference element, and changing the variable again. ⋄

Let $\pi_{cr} : H^1(\Omega) \to X_{cr}$ (resp. $\Pi_{cr} : \mathbf{H}^1(\Omega) \to \mathbf{X}_{cr}$) be the Crouzeix-Raviart interpolation operator:

$$\forall v \in H^1(\Omega), \quad \pi_{cr} v = \sum_{f \in \mathcal{I}_F} \underline{v}_f \, \psi_f, \tag{7.4.29}$$

$$\forall \mathbf{v} \in \mathbf{H}^1(\Omega), \quad \Pi_{cr}(\mathbf{v}) = (\pi_{cr}(v_i))_{i=1}^d. \tag{7.4.30}$$

The space of the discrete velocities of the Crouzeix-Raviart finite element is $\mathbf{X}_{0,cr}$. The space of the discrete pressure is $Q_{0,h}$, such that:

$$Q_{0,h} := \mathcal{P}^0_{disc}(\mathcal{T}_h) \cap L^2_{zmv}(\Omega) \tag{7.4.31}$$

Let $\mathbf{V}_{0,cr}$ be the subspace of $\mathbf{X}_{0,cr}$ functions which divergence is discretely vanishing:

$$\mathbf{V}_{0,cr} := \left\{ \mathbf{v}_h \in \mathbf{X}_{0,cr} \,|\, \forall q_{0,h} \in Q_{0,h}, \, (\mathrm{div}_h \, \mathbf{v}_h, q_{0,h})_{L^2(\Omega)} = 0 \right\}. \tag{7.4.32}$$

The discrete velocities in $\mathbf{V}_{0,cr}$ are piecewise divergence free, cf. Proposition 7.4.4. We have the tools to establish the following result.

Proposition 7.4.8. *Let $s \in [0,1]$. The Crouzeix-Raviart interpolation operator Π_{cr} can play the role of a Fortin operator for the $\mathbf{P}^1_{nc} \times P^0$ finite element:*

$$\forall \mathbf{v} \in \mathbf{H}^1(\Omega) \quad (\mathrm{div}_h \, \Pi_{cr} \mathbf{v}, q_h)_{L^2(\Omega)} = (\mathrm{div} \, \mathbf{v}, q_h)_{L^2(\Omega)}, \quad \forall q_h \in \mathcal{P}^0_{disc}(\mathcal{T}_h). \tag{7.4.33}$$

$$\forall \mathbf{v} \in \mathbf{H}^{s+1}(\Omega) \qquad \|\mathbf{v} - \Pi_{cr} \mathbf{v}\|_h \leq 2 \, \sigma h^s |\mathbf{v}|_{\mathbf{H}^{s+1}(\Omega)}. \tag{7.4.34}$$

Moreover, we have the following estimates:

$$\forall v \in H^1(\Omega), \qquad \|\Pi_{cr}(\mathbf{v})\|_h \leq \| \boldsymbol{\nabla} \mathbf{v}\|_{\mathbb{L}^2(\Omega)}. \tag{7.4.35}$$

$$\forall v \in H^{1+s}(\Omega), \forall \ell \in \mathcal{I}_K, \quad \|\mathbf{v} - \Pi_{cr} \mathbf{v}\|_{L^2(\partial K_\ell)} \leq 2 \, (\sigma_\ell \, h_\ell)^{\frac{1}{2}} \, (h_\ell)^s \, |\mathbf{v}|_{\mathbf{H}^{1+s}(K_\ell)}. \tag{7.4.36}$$

Proof. Notice that for all $v \in H^1(\Omega)$, $\pi_{cr} v_{|K_\ell} = \pi^\ell_{cr} v$. Then, we deduce (7.4.35) from (7.4.25), (7.4.34) from (7.4.27) and (7.4.36) from (7.4.28). By definition of Π_{cr} and using an integration by parts, we obtain that for all $\ell \in \mathcal{I}_K$: $(\mathrm{div}(\Pi_{cr}\mathbf{v} - \mathbf{v}), q_h)_{L^2(K_\ell)} = \sum_{f \in \mathcal{I}_{f,\ell}} -q_{h|K_\ell} \int_{F_f} (\Pi_{cr}\mathbf{v} - \mathbf{v}) \cdot \mathbf{n}_f = 0$. We deduce (7.4.33) by summation. ⋄

We will need the following Lemma in §7.5.1. The construction of $\mathbf{v}_{q_{0,h}}$ is explained in the proof of Proposition 7.4.5, letting $\mathbf{v}_{q_{0,h}} = \mathbf{v}_{h,p'_h}$ and $C_{nc} = 1$.

Lemma 7.4.1. *For all $q_{0,h} \in Q_{0,h}$, it exists $\mathbf{v}_{q_{0,h}} \in \mathbf{X}_{0,cr}$ such that:*

$$-(\mathrm{div}_h \, \mathbf{v}_{q_{0,h}}, q_{0,h})_{L^2(\Omega)} = \nu^{-1} \, \|q_{0,h}\|^2_{L^2(\Omega)}, \tag{7.4.37}$$

$$\|\mathbf{v}_{q_{0,h}}\|_h \leq \nu^{-1} \, C_{\mathrm{div}} \, \|q_{0,h}\|_{L^2(\Omega)}. \tag{7.4.38}$$

Using estimates (7.4.35) and (7.4.4), we apply the proof of Theorem 7.4.1 to state the

Theorem 7.4.3. *Let $\mathcal{X}_h = \mathcal{X}_{cr}$. Then the continuous bilinear form $a_h(\cdot,\cdot)$ is T-coercive and Problem (7.4.10) is well-posed.*

Here $C^{nc}_{\min} = \min((C_{\mathrm{div}})^2, 1) = 1$ and $C^{nc}_{\max} = C_{\max}$: the stability constant of the nonconforming Crouzeix-Raviart mixed finite elements is independent of the mesh. This is not the case for higher order (see [81, Theorem 2.2]).

For higher order, we cannot built the interpolation operator component by component, since higher-order divergence moments must be preserved. Thus, for $k > 1$, we must build Π_{nc} so that for all $\mathbf{v} \in \mathbf{H}^1(\Omega)$, for all $\ell \in \mathcal{I}_K$, for all $q \in P^{k-1}(K_\ell)$: $\displaystyle\int_{K_\ell} q \, \mathrm{div} \, \Pi_{nc} \mathbf{v} = \int_{K_\ell} q \, \mathrm{div} \, \mathbf{v}$. We recall that by

integration by parts, we have:

$$\int_{K_\ell} q \, \mathrm{div}\, \Pi_{nc}\mathbf{v} + \int_{K_\ell} \nabla q \cdot \Pi_{nc}\mathbf{v} = \int_{\partial K_\ell} q \, \Pi_{nc}\mathbf{v} \cdot \mathbf{n}_{|\partial K_\ell}. \tag{7.4.39}$$

Hence, to obtain a local estimate of $\| \boldsymbol{\nabla} \Pi_{nc}\mathbf{v}\|_{\mathbb{L}^2(K_\ell)}$, we will need the following lemma (proven using an integration by parts and Cauchy-Schwarz):

Lemma 7.4.2. *Let $\mathbf{v} \in \mathbf{H}^1(K_\ell)$ and $q \in P^{k-1}(K_\ell) \cap L^2_{zmv}(K_\ell)$. We have:*

$$\left| \int_{\partial K_\ell} q \, (\mathbf{v} - \underline{\mathbf{v}}_\ell) \cdot \mathbf{n}_{|\partial K_\ell} \right| \lesssim h_\ell \, \| \nabla q\|_{\mathbf{L}^2(K_\ell)} \, \| \boldsymbol{\nabla} \mathbf{v}\|_{\mathbb{L}^2(K_\ell)} \tag{7.4.40}$$

Proof. We have by integration by parts, and then using Cauchy-Schwarz inequality:

$$\begin{aligned}
\left| \int_{\partial K_\ell} q \, (\mathbf{v} - \underline{\mathbf{v}}_\ell) \cdot \mathbf{n}_{|\partial K_\ell} \right| &\leq \left| \int_{K_\ell} q \ \mathrm{div}(\mathbf{v} - \underline{\mathbf{v}}_\ell) \right| + \left| \int_{K_\ell} \nabla q \cdot (\mathbf{v} - \underline{\mathbf{v}}_\ell) \right|, \\
&\leq \sqrt{d} \, \|q\|_{L^2(K_\ell)} \, \| \boldsymbol{\nabla} (\mathbf{v} - \underline{\mathbf{v}}_\ell)\|_{\mathbb{L}^2(K_\ell)} + \| \nabla q\|_{\mathbf{L}^2(K_\ell)} \, \|(\mathbf{v} - \underline{\mathbf{v}}_\ell)\|_{\mathbf{L}^2(K_\ell)}, \\
&\leq (\sqrt{d} + 1) \, \pi^{-1} \, h_\ell \, \| \nabla q\|_{\mathbf{L}^2(K_\ell)} \, \| \boldsymbol{\nabla} \mathbf{v}\|_{\mathbb{L}^2(K_\ell)}, \text{ using (7.3.7) twice.}
\end{aligned}$$

◇

In the next Section, we will see that for $k = 2$, $d = 2$, we will need Lemma 7.4.2. For $k \geq 3$, it could be necessary to bound the tangential components of $\mathbf{v} - \underline{\mathbf{v}}_\ell$. To do so, we would need to preserve curl integrals on K_ℓ. Indeed, by integration by parts, we have:

$$\begin{aligned}
\text{For } d = 2,\ \mathbf{v} \in \mathbf{H}^1(\Omega),\ q \in P^{k-1}(K_\ell): \quad & \int_{K_\ell} q \ (\mathbf{curl}\, q \cdot \mathbf{v} - \mathrm{curl}\, \mathbf{v}\, q) = \int_{\partial K_\ell} q \, \mathbf{v} \times \mathbf{n}_{|\partial K_\ell}. \\
\text{For } d = 3,\ \mathbf{v} \in \mathbf{H}^1(\Omega),\ \mathbf{w} \in \mathbf{P}^{k-1}(K_\ell): \quad & \int_{K_\ell} (\mathbf{w} \cdot \mathbf{curl}\, \mathbf{v} - \mathbf{curl}\, \mathbf{w} \cdot \mathbf{v}) \\
& \qquad = \int_{\partial K_\ell} (\mathbf{n}_{|\partial K_\ell} \times \mathbf{v} \times \mathbf{n}_{|\partial K_\ell}) \cdot (\mathbf{w} \times \mathbf{n}_{|\partial K_\ell}).
\end{aligned}$$

7.4.2 The $\mathbf{P}^2_{nc} \times P^1_{disc}$ finite element

We consider here the case $k = 2$, $d = 2$ and we study the so-called Fortin-Soulie mixed finite elements [152]. We consider a shape-regular triangulation sequence $(\mathcal{T}_h)_h$.

Let us consider X_{fs} (resp. $X_{0,fs}$), the space of nonconforming approximation of $H^1(\Omega)$ (resp. $H^1_0(\Omega)$) of order 2:

$$\begin{aligned}
X_{fs} &= \left\{ v_h \in P^2_{disc}(\mathcal{T}_h) \, ; \quad \forall f \in \mathcal{I}^i_F, \, \forall q_h \in P^1(F_f), \, \int_{F_f} [v_h]_{F_f} \, q_h = 0 \right\} ; \\
X_{0,fs} &= \left\{ v_h \in X_{fs} \, ; \quad \forall f \in \mathcal{I}^b_F, \, \forall q_h \in P^1(F_f), \, \int_{F_f} v_h \, q_h = 0 \right\}.
\end{aligned} \tag{7.4.41}$$

The space of nonconforming approximation of $\mathbf{H}^1(\Omega)$ (resp. $\mathbf{H}^1_0(\Omega)$) of order 2 is $\mathbf{X}_{fs} = (X_{fs})^2$ (resp. $\mathbf{X}_{0,fs} = (X_{0,fs})^2$). We set $\mathcal{X}_{fs} = \mathbf{X}_{0,fs} \times Q_{fs}$ where $Q_{fs} := P^1_{disc}(\mathcal{T}_h) \cap L^2_{zmv}(\Omega)$.
The building of a basis for $X_{0,fs}$ is more involved than for $X_{0,cr}$ since we cannot use two points per facet as degrees of freedom. Indeed, for all $\ell \in K_\ell$, it exists a polynomial of order 2 vanishing on the Gauss-Legendre points of the facets of the boundary ∂K_ℓ. Let $f \in \mathcal{I}_F$. The barycentric coordinates of the two Gauss-Legendre points $(p_{+,f}, p_{-,f})$ on F_f are such that:

$$p_{+,f} = (c_+, c_-), \ p_{-,f} = (c_-, c_+), \ \text{where } c_\pm = (1 \pm 1/\sqrt{3})/2.$$

These points can be used to integrate exactly order three polynomials:

$$\forall g \in P^3(F_f), \int_{F_f} g = \frac{|F_f|}{2} \left(g(p_{+,f}) + g(p_{-,f}) \right).$$

For all $\ell \in \mathcal{I}_K$, we define the quadratic function ϕ_{K_ℓ} that vanishes on the six Gauss-Legendre points of the facets of K_ℓ (see Figure 7.1):

$$\phi_{K_\ell} := 2 - 3 \sum_{i \in \mathcal{I}_{S,\ell}} \lambda_{i,\ell}^2 \text{ such that } \quad \forall f \in \mathcal{I}_{F,\ell}, \forall q \in P^1(F_f), \quad \int_{F_f} \phi_{K_\ell}\, q = 0. \tag{7.4.42}$$

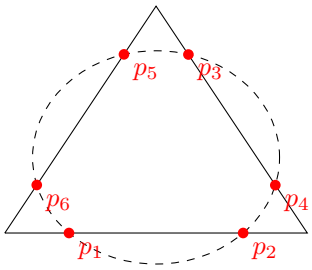


Figure 7.1: The six Gauss-Legendre points of an element K_ℓ and the elliptic function ϕ_{K_ℓ}.

We consider the set of the elliptic functions ϕ_{K_ℓ}:

$$\Phi_h := \{\phi_h \in L^2(\Omega); \quad \forall \ell \in \mathcal{I}_K,\, \phi_{h|K_\ell} = v_{K_\ell} \phi_{K_\ell},\, v_{K_\ell} \in \mathbb{R}\}. \tag{7.4.43}$$

We also define the spaces of P^2-Lagrange functions:

$$X_{LG} := \left\{ v_h \in H^1(\Omega); \quad \forall \ell \in \mathcal{I}_K,\, v_{h|K_\ell} \in P^2(K_\ell) \right\}, \quad X_{0,LG} := \left\{ v_h \in X_{LG}; \quad v_{h|\partial\Omega} = 0 \right\}.$$

The proposition below, cf. [152, Proposition 1], allows to build a basis for $X_{0,fs}$:

Proposition 7.4.9. *We have the following decomposition: $X_{fs} = X_{LG} + \Phi_h$ with $\dim(X_{LG} \cap \Phi_h) = 1$. Any function of X_{fs} can be written as the sum of a function of X_{LG} and a function of Φ_h. This representation can be made unique by specifying one degree of freedom.*

Notice that $\Phi_h \cap X_{LG} = \text{vect}(v_\Phi)$, where for all $\ell \in \mathcal{I}_K$, $v_{\Phi|K_\ell} = \phi_{K_\ell}$. Then, counting the degrees of freedom, one can show that $\dim(X_{fs}) = \dim(X_{LG}) + \dim(\Phi_h) + 1$. For problems involving Dirichlet boundary conditions we can prove thus that for $X_{0,fs}$ the representation is unique and $X_{0,fs} = X_{0,LG} \oplus \Phi_h$. We have $X_{LG} = \text{vect}\left((\phi_{S_i})_{i \in \mathcal{I}_S}, (\phi_{F_f})_{f \in \mathcal{I}_F} \right)$ where the basis functions are such that:

$$\forall i,j \in \mathcal{I}_S, \forall f,f' \in \mathcal{I}_F: \quad \phi_{S_i}(S_j) = \delta_{i,j},\, \phi_{S_i}(M_f) = 0, \quad \phi_{M_f}(M_{f'}) = \delta_{f,f'},\, \phi_{M_f}(S_i) = 0.$$

For all $\ell \in \mathcal{I}_K$, we will denote by $(\phi_{\ell,j})_{j=1}^6$ the local nodal basis such that:

$$(\phi_{\ell,j})_{j=1}^3 = (\phi_{S_i|K_\ell})_{i \in \mathcal{I}_{S,\ell}} \quad \text{and} \quad (\phi_{\ell,j})_{j=4}^6 = (\phi_{F_f|K_\ell})_{f \in \mathcal{I}_{F,\ell}}.$$

The spaces X_{fs} and $X_{0,fs}$ are such that:

$$\begin{aligned} X_{fs} &= \text{vect}\left((\phi_{S_i})_{i \in \mathcal{I}_S}, (\phi_{F_f})_{f \in \mathcal{I}_F}, (\phi_{K_\ell})_{\ell \in \mathcal{I}_K} \right), \\ X_{0,fs} &= \text{vect}\left((\phi_{S_i})_{i \in \mathcal{I}_S^i}, (\phi_{F_f})_{f \in \mathcal{I}_F^i}, (\phi_{K_\ell})_{\ell \in \mathcal{I}_K} \right). \end{aligned} \tag{7.4.44}$$

We propose here an alternative definition of the Fortin interpolation operator proposed in [152]. Let us first recall the Scott-Zhang interpolation operator [231, 93]. For all $i \in \mathcal{I}_S$, we choose some $\ell_i \in \mathcal{I}_{K,i}$,

and we build the $L^2(K_{\ell_i})$-dual basis $(\tilde{\phi}_{\ell_i,j})_{j=1}^6$ of the local nodal basis such that:

$$\forall j,\, j' \in \{1,\dots,6\}, \quad \int_{K_{\ell_i}} \tilde{\phi}_{\ell_i,j}\, \phi_{\ell_i,j'} = \delta_{j,j'}.$$

Let us define the Fortin-Soulie interpolation operator for scalar functions by:

$$\pi_{fs} : \left\{ \begin{array}{rcl} H^1(\Omega) & \to & X_{FS} \\ v & \mapsto & \tilde{\pi} v + \displaystyle\sum_{\ell \in \mathcal{I}_K} v_{K_\ell} \phi_{K_\ell} \end{array} \right. , \text{ with } \tilde{\pi} v = \sum_{i \in \mathcal{I}_S} v_{S_i} \phi_{S_i} + \sum_{f \in \mathcal{I}_F} v_{F_f} \phi_{F_f}. \tag{7.4.45}$$

- The coefficients $(v_{S_i})_{i\in\mathcal{I}_S}$ are fixed so that: $\forall i \in \mathcal{I}_S$, $v_{S_i} = \int_{K_{\ell,i}} v\, \tilde{\phi}_{\ell_i,j_i}$, where j_i is the index such that $\int_{K_{\ell_i}} \tilde{\phi}_{\ell_i,j_i}\, \phi_{S_i|K_{\ell_i}} = 1$.
- The coefficients $\left(v_{F_f}\right)_{f\in\mathcal{I}_F}$ are fixed so that: $\forall f \in \mathcal{I}_F$, $\int_{F_f} \tilde{\pi} v = \int_{F_f} v$.
- The coefficients v_{K_ℓ} are fixed so that: $\int_{K_\ell} \pi_{fs} v = \int_{K_\ell} v$.

The definition (7.4.45) is more general than the one given in [152], which holds for $v \in H^2(\Omega)$.

We set $\mathbf{v}_{S_i} := (\,\tilde{\pi} v_1(S_i), \tilde{\pi} v_2(S_i))^T$ and $\mathbf{v}_{F_f} := (\,\tilde{\pi} v_1(F_f), \tilde{\pi} v_2(F_f)\,)^T$.

We can define two different Fortin-Soulie interpolation operators for vector functions. First, let

$$\tilde{\Pi}_{fs} : \left\{ \begin{array}{rcl} \mathbf{H}^1(\Omega) & \to & \mathbf{X}_{fs} \\ \mathbf{v} & \mapsto & \displaystyle\sum_{i \in \mathcal{I}_S} \mathbf{v}_{S_i} \phi_{S_i} + \sum_{f \in \mathcal{I}_F} \mathbf{v}_{F_f} \phi_{F_f} + \sum_{\ell \in \mathcal{I}_K} \tilde{\mathbf{v}}_{K_\ell} \phi_{K_\ell}, \end{array} \right.$$

where the coefficients $(\tilde{\mathbf{v}}_{K_\ell})_{\ell\in\mathcal{I}_K}$ are such that:

$$\forall \ell \in \mathcal{I}_K, \quad \int_{K_\ell} \tilde{\Pi}_{fs} \mathbf{v} = \int_{K_\ell} \mathbf{v}. \tag{7.4.46}$$

The interpolation operator $\tilde{\Pi}_{fs}$ preserves the local averages, but it doesn't preserve the divergence. We then define a second interpolation operator which preserves the divergence in a weak sense:

$$\Pi_{fs} : \left\{ \begin{array}{rcl} \mathbf{H}^1(\Omega) & \to & \mathbf{X}_{fs} \\ \mathbf{v} & \mapsto & \displaystyle\sum_{i \in \mathcal{I}_S} \mathbf{v}_{S_i} \phi_{S_i} + \sum_{f \in \mathcal{I}_F} \mathbf{v}_{F_f} \phi_{F_f} + \sum_{\ell \in \mathcal{I}_K} \mathbf{v}_{K_\ell} \phi_{K_\ell} \end{array} \right. .$$

For all $\ell \in \mathcal{I}_K$, the vector coefficient $\mathbf{v}_{K_\ell} \in \mathbb{R}^2$ is now fixed so that condition (7.4.12) is satisfied. We can impose for example that the projection $\Pi_{fs}\mathbf{v}$ satisfies:

$$\int_{K_\ell} T_\ell^{-1}(\mathbf{x})\, \operatorname{div} \Pi_{fs} \mathbf{v} = \int_{K_\ell} T_\ell^{-1}(\mathbf{x})\, \operatorname{div} \mathbf{v}. \tag{7.4.47}$$

Notice that due to (7.4.42), the patch-test condition is still satisfied.

Proposition 7.4.10. *The Fortin-Soulie interpolation operator Π_{fs} is such for all $0 \le s \le 1$, for all $\mathbf{v} \in \mathbf{H}^{1+s}(\Omega)$, we have:*

$$\begin{aligned} \forall \ell \in \mathcal{I}_K, \quad \| \boldsymbol{\nabla}(\Pi_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)} &\lesssim (\sigma_\ell)^2\, (h_\ell)^s\, |\mathbf{v}|_{\mathbf{H}^{1+s}(K_\ell)}, & (7.4.48) \\ \|\Pi_{fs}\mathbf{v} - \mathbf{v}\|_h &\lesssim \sigma^2\, h^s\, |\mathbf{v}|_{\mathbf{H}^{1+s}(\Omega)}. & (7.4.49) \end{aligned}$$

Remark 7.4.3. *Albeit we are inspired by the proof of [121, Lemma 4], we changed the transition from Equation (4.27) to (4.29) there by using only the properties related to the normal component of the velocity, cf.* (7.4.40)*. In the original proof, one needs a stronger assumption on the regularity of* $\mathbf{v}$

(namely, $\mathbf{v} \in \bigcap_{0<s<s_\Omega} \mathbf{H}^{1+s}(\Omega)$ *with* $s_\Omega > \frac{1}{2}$*). Finally, because we do not split the integral over the boundaries of elements into the sum of* $d+1$ *integrals over the facets, we obtain purely local estimates, which appear to be new for the Fortin-Soulie element in the case of low-regularity fields* $\mathbf{v}$*.*

Proof. Let $\mathbf{v} \in \mathbf{H}^1(\Omega)$. We have:

$$\| \boldsymbol{\nabla} (\Pi_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)} \leq \| \boldsymbol{\nabla} (\Pi_{fs}\mathbf{v} - \tilde{\Pi}_{fs}\mathbf{v})\|_{\mathbb{L}^2(K_\ell)} + \| \boldsymbol{\nabla} (\tilde{\Pi}_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)}.$$

Notice that for all $\ell \in \mathcal{I}_K$, $(\Pi_{fs}\mathbf{v} - \tilde{\Pi}_{fs}\mathbf{v})_{|K_\ell} = (\mathbf{v}_{K_\ell} - \tilde{\mathbf{v}}_{K_\ell})\,\phi_{K_\ell}$. Using (7.3.5), we obtain that:

$$\| \boldsymbol{\nabla} (\Pi_{fs}\mathbf{v} - \tilde{\Pi}_{fs}\mathbf{v})\|_{\mathbb{L}^2(K_\ell)} \lesssim |\mathbf{v}_{K_\ell} - \tilde{\mathbf{v}}_{K_\ell}|\, \| \nabla\, \phi_{K_\ell}\|_{\mathbf{L}^2(K_\ell)} \lesssim \|{\mathbb{B}_\ell}^{-1}\|\, |K_\ell|^{\frac{1}{2}}\, |\mathbf{v}_{K_\ell} - \tilde{\mathbf{v}}_{K_\ell}|. \tag{7.4.50}$$

Let us estimate $|\mathbf{v}_{K_\ell} - \tilde{\mathbf{v}}_{K_\ell}|$. On the one hand, we have[2]:

$$\begin{aligned}
\int_{K_\ell} (\Pi_{fs}\mathbf{v} - \tilde{\Pi}_{fs}\mathbf{v}) &= \int_{K_\ell} (\Pi_{fs}\mathbf{v} - \mathbf{v}) \text{ from (7.4.46)},\\
&= \int_{\partial K_\ell} \mathbf{x}\, (\Pi_{fs}\mathbf{v} - \mathbf{v}) \cdot \mathbf{n}_{|\partial K_\ell} \text{ by IBP and using (7.4.47)},\\
&= \int_{\partial K_\ell} (\mathbf{x} - \underline{\mathbf{x}})\, (\Pi_{fs}\mathbf{v} - \mathbf{v}) \cdot \mathbf{n}_{|\partial K_\ell} \text{ since } \int_{\partial K_\ell} (\Pi_{fs}\mathbf{v} - \mathbf{v}) \cdot \mathbf{n}_{|\partial K_\ell} = 0,\\
&= \int_{\partial K_\ell} (\mathbf{x} - \underline{\mathbf{x}})\, (\tilde{\Pi}_{fs}\mathbf{v} - \mathbf{v}) \cdot \mathbf{n}_{|\partial K_\ell} \text{ from (7.4.42)}.
\end{aligned}$$

On the other hand, it holds: $\int_{K_\ell} (\Pi_{fs}\mathbf{v} - \tilde{\Pi}_{fs}\mathbf{v}) = (\mathbf{v}_{K_\ell} - \tilde{\mathbf{v}}_{K_\ell}) \int_{K_\ell} \phi_{K_\ell} = \frac{|K_\ell|}{4}\, (\mathbf{v}_{K_\ell} - \tilde{\mathbf{v}}_{K_\ell})$. Hence we have:

$$|\mathbf{v}_{K_\ell} - \tilde{\mathbf{v}}_{K_\ell}| \leq 4\, |K_\ell|^{-1} \left| \int_{\partial K_\ell} (\mathbf{x} - \underline{\mathbf{x}})\, (\tilde{\Pi}_{fs}\mathbf{v} - \mathbf{v}) \cdot \mathbf{n}_{|\partial K_\ell} \right|. \tag{7.4.51}$$

In order to bound the right-hand-side of (7.4.51) component by component, we can use Lemma 7.4.2, with $q = x_{d'} - \int_{K_\ell} x_{d'}/|K_\ell|$ $(d' = 1, 2)$, so that $\| \nabla\, q\|_{\mathbf{L}^2(K_\ell)} = |K_\ell|^{\frac{1}{2}}$. We obtain:

$$|\mathbf{v}_{K_\ell} - \tilde{\mathbf{v}}_{K_\ell}| \leq 4\, d \times (\sqrt{d} + 1)\, |K_\ell|^{-\frac{1}{2}}\, \pi^{-1}\, h_\ell\, \| \boldsymbol{\nabla} (\tilde{\Pi}_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)}. \tag{7.4.52}$$

Combining (7.4.50) and (7.4.52), we have:

$$\| \boldsymbol{\nabla} (\Pi_{fs}\mathbf{v} - \tilde{\Pi}_{fs}\mathbf{v})\|_{\mathbb{L}^2(K_\ell)} \lesssim \|{\mathbb{B}_\ell}^{-1}\|\, h_\ell\, \| \boldsymbol{\nabla} (\tilde{\Pi}_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)} \lesssim \sigma_\ell\, \| \boldsymbol{\nabla} (\tilde{\Pi}_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)}.$$

For all $\mathbf{v} \in \mathbf{P}^2(K_\ell)$ we have $\tilde{\Pi}_{fs}(\mathbf{v}) = \mathbf{v}$ and $\hat{\tilde{\Pi}}_{fs}\hat{\mathbf{v}}_\ell = \hat{\mathbf{v}}_\ell$. Hence, using Bramble-Hilbert/Deny-Lions Lemma [147, Lemma 11.9], we have for $m = 0, 1$, for all $\mathbf{v} \in \mathbf{H}^{1+m}(\Omega)$, for all $\ell \in \mathcal{I}_K$, $\| \boldsymbol{\nabla} (\tilde{\Pi}_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)} \lesssim \sigma_\ell\, (h_\ell)^m\, |\mathbf{v}|_{\mathbf{H}^{1+m}(K_\ell)}$. We deduce that for $m = 0, 1$, for all $\mathbf{v} \in \mathbf{H}^{1+m}(\Omega)$, for all $\ell \in \mathcal{I}_K$, $\| \boldsymbol{\nabla} (\Pi_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)} \lesssim (\sigma_\ell)^2\, (h_\ell)^m\, |\mathbf{v}|_{\mathbf{H}^{1+m}(K_\ell)}$ and by summation $\| \boldsymbol{\nabla} (\Pi_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(\Omega)} \lesssim \sigma^2\, h^m\, |\mathbf{v}|_{\mathbf{H}^{1+m}(\Omega)}$. Using interpolation property [237, Lemma 22.2], we obtain (7.4.48) and (7.4.49). ⋄

Hence, using the triangular inequality, we have:

$$\| \boldsymbol{\nabla}\, \Pi_{fs}\mathbf{v}\|_{\mathbb{L}^2(K_\ell)} \leq \| \boldsymbol{\nabla} (\Pi_{fs}\mathbf{v} - \mathbf{v})\|_{\mathbb{L}^2(K_\ell)} + \| \boldsymbol{\nabla}\, \mathbf{v}\|_{\mathbb{L}^2(K_\ell)} \lesssim ((\sigma_\ell)^2 + 1)\|\mathbf{v}\|_{\mathbf{H}^1(K_\ell)}.$$

By summation over ℓ, we deduce that the coefficient C_{nc} in (7.4.11) is here equal $\sigma^2 + 1$. We recall that the discrete Poincaré–Steklov inequality (7.4.4) holds.

Theorem 7.4.4. *Let* $\mathcal{X}_h = \mathcal{X}_{fs}$*. Then the continuous bilinear form* $a_h(\cdot, \cdot)$ *is* T*-coercive and Problem* (7.4.10) *is well-posed.*

Proof. According to Proposition 7.4.10, the operator Π_{fs} satisfies (7.4.11)-(7.4.12), so that we apply Theorem 7.4.1. ⋄

Notice that in the recent paper [228], the inf-sup condition of the mixed Fortin-Soulie finite element

[2] Let set $\mathbf{w} = \Pi_{fs}\mathbf{v} - \mathbf{v} = (w_1, w_2)^T$ and $\mathbf{x} := (x_1, x_2)^T$. By IBP, we have for $d' = 1, 2$: $\int_{\partial K_\ell} x_{d'}\, \mathbf{w} \cdot \mathbf{n}_{|\partial K_\ell} = \int_{K_\ell} x_{d'}\, \operatorname{div} \mathbf{w} + \int_{K_\ell} \mathbf{w} \cdot \mathbf{e}_{d'}$. Due to (7.4.47), $\int_{K_\ell} x_{d'}\, \operatorname{div} \mathbf{w} = 0$, so that for $d' = 1, 2$: $\int_{K_\ell} w_{d'} = \int_{\partial K_\ell} x_{d'}\, \mathbf{w} \cdot \mathbf{n}_{|\partial K_\ell}$.

is proven directly on a triangle and then using the macro-element technique [234], but it seems difficult to use this technique to build a Fortin operator, which is needed to compute error estimates.

The study can be extended to higher orders for $d = 2$ using the following papers: [21] for $k \geq 4$, k even, [82] for $k = 3$ and [81] for $k \geq 5$, k odd. In [131], the authors propose a local Fortin operator for the lowest order Taylor-Hood finite element [239] for $d = 3$.

7.4.3 Improving consistency

Consider Problem (7.2.1) with data $\mathbf{f} = -\nabla \phi$, where $\phi \in H^1(\Omega) \cap L^2_{zmv}(\Omega)$. The unique solution is then $(\mathbf{u}, p) := (0, \phi)$. By integrating by parts, the source term in (7.2.2) reads:

$$\forall \mathbf{v} \in \mathbf{H}^1_0(\Omega), \quad \int_\Omega \mathbf{f} \cdot \mathbf{v} = \int_\Omega \phi \ \mathrm{div}\, \mathbf{v}. \tag{7.4.53}$$

Recall that the nonconforming space $\mathbf{X}_h$ defined in (7.4.1) is a subset of $\mathcal{P}_h\mathbf{H}^1$: using a nonconforming finite element method, the integration by parts must be done on each element of the triangulation, and we have:

$$\forall \mathbf{v} \in \mathcal{P}_h\mathbf{H}^1, \quad \int_\Omega \mathbf{f} \cdot \mathbf{v} = (\mathrm{div}_h\, \mathbf{v}, \phi)_{L^2(\Omega)} + \sum_{f \in \mathcal{I}_F} \int_{F_f} [\mathbf{v}]_{F_f} \cdot \mathbf{n}_f \, \phi. \tag{7.4.54}$$

Using Lemma 7.4.2, we have: $\displaystyle\sum_{f \in \mathcal{I}_F} \int_{F_f} [\mathbf{v}]_{F_f} \cdot \mathbf{n}_f \, \phi \lesssim h \, \|\mathbf{v}_h\|_h \, \| \nabla \phi\|_{\mathbf{L}^2(\Omega)}$. Applying (7.4.54) to the right-hand-side of (7.4.10) and choosing $\mathbf{v}_h = \mathbf{u}_h$, it holds: $\nu \, \|\mathbf{u}_h\|_h \lesssim h \| \nabla \phi\|_{\mathbf{L}^2(\Omega)}$ (as expected by (7.4.17)). Hence, the term with the jumps acts as a numerical source for the discrete velocity, which numerical influence is proportional to h/ν. Thus, we cannot obtain exactly $\mathbf{u}_h = 0$.

Linke proposed in [194] to project the test function $\mathbf{v}_h \in \mathbf{X}_h$ on a discrete subspace of $\mathbf{H}(\mathrm{div};\, \Omega)$, like Raviart-Thomas or Brezzi-Douglas-Marini finite elements (see [213, 71], or the monograph [50]). Let $\Pi_{\mathrm{div}} : \mathbf{X}_{0,h} \to \mathcal{P}^k_{disc}(\mathcal{T}_h) \cap \mathbf{H}_0(\mathrm{div}; \Omega)$ be some interpolation operator built so that for all $\mathbf{v}_h \in \mathbf{X}_{0,h}$, for all $\ell \in \mathcal{I}_K$, $(\mathrm{div}\, \Pi_{\mathrm{div}} \mathbf{v}_h)_{|K_\ell} = \mathrm{div}\, \mathbf{v}_{h|K_\ell}$. Integrating by parts, we have for all $\mathbf{v}_h \in \mathbf{X}_{0,h}$:

$$\int_\Omega \mathbf{f} \cdot \Pi_{\mathrm{div}} \mathbf{v}_h = \int_\Omega \phi \ \mathrm{div}\, \Pi_{\mathrm{div}} \mathbf{v}_h = \sum_{\ell \in K_\ell} \int_{K_\ell} \phi \ \mathrm{div}\, \Pi_{\mathrm{div}} \mathbf{v}_h = \sum_{\ell \in K_\ell} \int_{K_\ell} \phi \ \mathrm{div}\, \mathbf{v}_h = (\mathrm{div}_h\, \mathbf{v}_h, \phi)_{L^2(\Omega)}.$$

The projection Π_{div} allows to eliminate the terms of the integrals of the jumps in (7.4.54).
Let us write Problem (7.4.10) as: Find $(\mathbf{u}_h, p_h) \in \mathcal{X}_h$ such that

$$a_h \left((\mathbf{u}_h, p_h), (\mathbf{v}_h, q_h) \right) = \ell_{\mathbf{f}} \left(\, (\Pi_{\mathrm{div}} \mathbf{v}_h, q_h) \, \right) \quad \forall (\mathbf{v}_h, q_h) \in \mathcal{X}_h. \tag{7.4.55}$$

In the case of $\mathcal{X}_h = \mathcal{X}_{cr}$ and a projection on Brezzi-Douglas-Marini finite elements, the following error estimate holds if $(\mathbf{u}, p) \in \mathbf{H}^2(\Omega) \times H^1(\Omega)$:

$$\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} \leq \widetilde{C} \, h^2 \, |\mathbf{u}|_{\mathbf{H}^2(\Omega)}, \tag{7.4.56}$$

where the constant $\widetilde{C}$ if independent of h. The proof is detailed in [66] for shape-regular meshes and [12] for anisotropic meshes. We remark that the error doesn't depend on the norm of the pressure nor on the ν parameter. We will provide some numerical results to illustrate the effectiveness of this formulation, even with a projection on the Raviart-Thomas finite elements, which, for a fixed polynomial order, are less precise than the Brezzi-Douglas-Marini finite elements.

For $k \in \mathbb{N}^\star$, the space of Raviart-Thomas finite elements can be defined as:

$$\mathbf{X}_{RT_k} := \left\{ \mathbf{v} \in \mathbf{H}(\mathrm{div};\, \Omega);\, \forall \ell \in \mathcal{I}_k,\, \mathbf{v}_{|K_\ell} = \mathbf{a}_\ell + b_\ell \mathbf{x} \,|\, (\mathbf{a}_\ell, b_\ell) \in P^k(K_\ell)^d \times P^k_H(K_\ell) \right\}.$$

Let $k \leq 1$. The Raviart–Thomas interpolation operator $\Pi_{RT_k} : \mathbf{H}^1(\Omega) \cup \mathbf{X}_h \to \mathbf{X}_{RT_k}$ is defined by: $\forall \mathbf{v} \in \mathbf{H}^1(\Omega) \cup \mathbf{X}_h$,

$$\left\{ \begin{array}{rl} \forall f \in \mathcal{I}_F, & \displaystyle\int_{F_f} \Pi_{RT_k}\mathbf{v} \cdot \mathbf{n}_f \, q = \int_{F_f} \mathbf{v} \cdot \mathbf{n}_f \, q, \quad \forall q \in P^k(F_f) \\ \text{for } k=1, \forall \ell \in \mathcal{I}_K, & \displaystyle\int_{K_\ell} \Pi_{RT_1}\mathbf{v} = \int_{K_\ell} \mathbf{v} \end{array} \right. . \tag{7.4.57}$$

Note that the Raviart–Thomas interpolation operator preserves the constants. Let $\mathbf{v}_h \in \mathbf{X}_h$. In order to compute the left-hand-side of (7.4.54), we must evaluate $(\Pi_{RT_k}\mathbf{v}_h)_{|K_\ell}$ for all $\ell \in \mathcal{I}_K$. Calculations are performed using the proposition below, which corresponds to [158, Lemma 3.11]:

Proposition 7.4.11. *Let $k \leq 1$. Let $\hat{\Pi}_{RT_k} : \mathbf{H}^1(\hat{K}) \to \mathbf{P}^k(\hat{K})$ be the Raviart–Thomas interpolation operator restricted to the reference element, so that: $\forall \hat{\mathbf{v}} \in \mathbf{H}^1(\hat{K})$,*

$$\left\{ \begin{array}{rl} \forall \hat{F} \in \partial\hat{K}, & \displaystyle\int_{\hat{F}} \hat{\Pi}_{RT_k}\hat{\mathbf{v}} \cdot \mathbf{n}_{\hat{F}} \, \hat{q} = \int_{\hat{F}} \hat{\mathbf{v}} \cdot \mathbf{n}_{\hat{F}} \, \hat{q}, \quad \forall \hat{q} \in P^k(\hat{F}) \\ \text{for } k=1, & \displaystyle\int_{\hat{K}} \hat{\Pi}_{RT_k}\hat{\mathbf{v}} = \int_{\hat{K}} \hat{\mathbf{v}} \end{array} \right. . \tag{7.4.58}$$

We then have: $\forall \ell \in \mathcal{I}_K$,

$$(\Pi_{RT_k}\mathbf{v})_{|K_\ell}(\mathbf{x}) = \mathbb{B}_\ell \left(\hat{\Pi}_{RT_k} {\mathbb{B}_\ell}^{-1} \hat{\mathbf{v}}_\ell \right) \circ {T_\ell}^{-1}(\mathbf{x}) \;\; \textit{where} \;\; \hat{\mathbf{v}}_\ell = \mathbf{v} \circ T_\ell(\hat{\mathbf{x}}). \tag{7.4.59}$$

The proof is based on the equality of the $\hat{F}$ and $\hat{K}$-moments of functions

$$(\Pi_{RT_k}\mathbf{v})_{|K_\ell} \circ T_\ell(\hat{\mathbf{x}}) \text{ and } \mathbb{B}_\ell \left(\hat{\Pi}_{RT_k} {\mathbb{B}_\ell}^{-1} \hat{\mathbf{v}}_\ell \right)(\hat{\mathbf{x}}).$$

For $k=0$, setting for $d' \in \{1, \ldots, d\}$: $\boldsymbol{\psi}_{f,d'} := \psi_f \, \mathbf{e}_{d'}$, we obtain that:

$$\forall \ell \in \mathcal{I}_K \, , \, \forall f \in \mathcal{I}_{F,\ell} \, , \quad (\Pi_{RT_0}\boldsymbol{\psi}_{f,d'})_{|K_\ell} = (d\,|K_\ell|)^{-1} \left(\mathbf{x} - \vec{OS}_{f,\ell} \right) \mathcal{S}_{f,\ell} \cdot \mathbf{e}_{d'} \, , \tag{7.4.60}$$

where $S_{f,\ell}$ is the vertex opposite to F_f in K_ℓ.

For $k=1$, the vector $(\Pi_{RT_1}\mathbf{v}_h)_{|K_\ell}$ is described by eight unknowns:

$$(\Pi_{RT_1}\mathbf{v}_h)_{|K_\ell} = \mathbb{A}_\ell \, \mathbf{x} + (\mathbf{b}_\ell \cdot \mathbf{x}) \, \mathbf{x} + \mathbf{d}_\ell, \text{ where } \mathbb{A}_\ell \in \mathbb{R}^{2\times 2}, \, \mathbf{b}_\ell \in \mathbb{R}^2, \, \mathbf{d}_\ell \in \mathbb{R}^2.$$

We compute only once the inverse of the matrix of the linear system (7.4.58), in $\mathbb{R}^{8\times 8}$.

In the Tables 7.1, 7.2 and 7.3, we report the L^2 velocity error $\varepsilon_0^\nu(\mathbf{u}_h)$ (cf. definition 7.4.1), where $\mathbf{u}_h$ is the solution to Problem (7.4.10) (columns $\mathbf{X}_{cr}$ and $\mathbf{X}_{fs}$) or (7.4.55) (columns $\mathbf{X}_{cr} + \Pi_{RT_0}$ and $\mathbf{X}_{fs} + \Pi_{RT_1}$) and h is the mesh size.

We first consider Stokes problem (7.2.1) in $\Omega = (0,1)^2$ with $\mathbf{u} = 0$, $p = (x_1)^3 + (x_2)^3 - 0.5$ so that $\mathbf{f} = \nabla p = 3 \left((x_1)^2, (x_2)^2 \right)^T$. We report in Table 7.1 $\varepsilon_0^\nu(\mathbf{u}_h)$ for $h = 5.00 \times 10^{-2}$ and different ν values.

ν	$\mathbf{X}_{cr}$	$\mathbf{X}_{cr} + \Pi_{RT_0}$	$\mathbf{X}_{fs}$	$\mathbf{X}_{fs} + \Pi_{RT_1}$
1.00×10^{-4}	7.96×10^{-4}	4.59×10^{-17}	8.81×10^{-7}	1.54×10^{-16}
1.00×10^{-5}	7.96×10^{-4}	4.59×10^{-17}	8.81×10^{-7}	1.54×10^{-16}
1.00×10^{-6}	7.96×10^{-4}	4.59×10^{-17}	8.81×10^{-7}	1.54×10^{-16}

Table 7.1: Values of $\varepsilon_0^\nu(\mathbf{u}_h)$ for $h = 5.00 \times 10^{-2}$

Here we have $\|(\mathbf{u}, p)\|_{\mathcal{X}} = \nu \|p\|_{L^2(\Omega)}$. Hence, the $\mathbf{L}^2(\Omega)$-norm of the discrete velocity $\|\mathbf{u}_h\|_{\mathbf{L}^2(\Omega)}$ is proportional to ν^{-1}. Using the projection, we obtain $\varepsilon_0^\nu(\mathbf{u}_h) = 0$ close to machine precision.

We now consider Stokes problem (7.2.1) in $\Omega = (0,1)^2$ with:

$$\mathbf{u} = \begin{pmatrix} (1-\cos(2\,\pi\,x_1))\,\sin(2\,\pi\,x_2) \\ (\cos(2\,\pi\,x_2)-1)\,\sin(2\,\pi\,x_1) \end{pmatrix},\ p = \sin(2\,\pi\,x_1) \text{ and } \mathbf{f} = -\nu\,\Delta\mathbf{u} + \nabla\,p.$$

We report in Table 7.2 (resp. 7.3) the values of $\varepsilon_0^\nu(\mathbf{u}_h)$ in the case $\nu = 1.00\times10^{-3}$ (resp. $\nu = 1.00\times10^{-4}$) for mesh sizes. We observe that when there is no projection, $\varepsilon_0^\nu(\mathbf{u}_h)$ is independent of ν, whereas using the projection, $\varepsilon_0^\nu(\mathbf{u}_h)$ is proportional to ν.

h	$\mathbf{X}_{cr}$	$\mathbf{X}_{cr}+\Pi_{RT_0}$	$\mathbf{X}_{fs}$	$\mathbf{X}_{fs}+\Pi_{RT_1}$
5.00×10^{-2}	1.32×10^{-3}	2.74×10^{-5}	4.73×10^{-6}	5.05×10^{-7}
2.50×10^{-2}	3.30×10^{-4}	6.93×10^{-6}	5.06×10^{-7}	6.42×10^{-8}
1.25×10^{-2}	8.25×10^{-5}	1.74×10^{-6}	6.31×10^{-8}	8.10×10^{-9}
6.25×10^{-3}	2.04×10^{-5}	4.35×10^{-7}	7.44×10^{-9}	1.03×10^{-9}
Rate	$h^{2.00}$	$h^{1.99}$	$h^{3.08}$	$h^{2.97}$

Table 7.2: Values of $\varepsilon_0^\nu(\mathbf{u}_h)$ for $\nu = 1.00\times10^{-3}$

h	$\mathbf{X}_{cr}$	$\mathbf{X}_{cr}+\Pi_{RT_0}$	$\mathbf{X}_{fs}$	$\mathbf{X}_{fs}+\Pi_{RT_1}$
5.00×10^{-2}	1.32×10^{-3}	2.74×10^{-6}	4.70×10^{-6}	5.05×10^{-8}
2.50×10^{-2}	3.30×10^{-4}	6.93×10^{-7}	5.10×10^{-7}	6.43×10^{-9}
1.25×10^{-2}	8.25×10^{-5}	1.74×10^{-7}	6.37×10^{-8}	8.11×10^{-10}
6.25×10^{-3}	2.04×10^{-5}	4.36×10^{-8}	7.51×10^{-9}	9.77×10^{-11}
Rate	$h^{2.00}$	$h^{1.99}$	$h^{3.08}$	$h^{2.99}$

Table 7.3: Values of $\varepsilon_0^\nu(\mathbf{u}_h)$ for $\nu = 1.00\times10^{-4}$

Let us consider Stokes problem (7.2.1) with a low-regular velocity. Let $\Omega = (0,1)^2$, $S_0 = (0.5, 0.5)$, and (r,θ) be the polar coordinates centred on S_0. We set:

$$\mathbf{u} = r^\alpha\mathbf{e}_\theta,\ p = r \text{ so that } \mathbf{f} := -\nu\,\boldsymbol{\Delta}\,\mathbf{u} + \nabla\,p = \nu\,(1-\alpha^2)\,r^{\alpha-2}\,\mathbf{e}_\theta + \mathbf{e}_r.$$

We report in Table 7.4 the values of $\varepsilon_0^\nu(\mathbf{u}_h)$ for $\nu = 1.00\times10^{-4}$, and for different for mesh sizes, with $\alpha = 1$ and $\alpha = 0.49$. For $\alpha = 1$, $\mathbf{u} = (-y,x)^T \in \mathbf{H}^2(\Omega)$. For $\alpha = 0.49$, $\mathbf{u} \in \bigcap_{0<s<\alpha} \mathbf{H}^{1+s}(\Omega)$, hence $\mathbf{u} \notin \mathbf{H}^2(\Omega)$. It seems that the Raviart-Thomas projection is less efficient in that last case.
In order to enhance the numerical results, one could also use a posteriori error estimators to adapt the

	$\alpha = 1$		$\alpha = 0.45$	
h	$\mathbf{X}_{fs}$	$\mathbf{X}_{fs}+\Pi_{RT_1}$	$\mathbf{X}_{fs}$	$\mathbf{X}_{fs}+\Pi_{RT_1}$
1.00×10^{-1}	3.03×10^{-5}	2.81×10^{-6}	3.05×10^{-5}	3.94×10^{-6}
5.00×10^{-2}	4.34×10^{-6}	1.54×10^{-6}	4.57×10^{-6}	2.15×10^{-6}
2.50×10^{-2}	4.72×10^{-7}	2.41×10^{-8}	9.70×10^{-7}	8.52×10^{-7}
Rate	$h^{3.00}$	$h^{3.43}$	$h^{2.48}$	$h^{1.11}$

Table 7.4: Values of $\varepsilon_0^\nu(\mathbf{u}_h)$, regular and low-regular velocity, $\nu = 1.00\times10^{-4}$.

mesh near point S_0 (see [123, 135] for $k = 1$ and [2] for $k = 2$).

Alternatively, using the nonconforming Crouzeix-Raviart mixed finite element method, one can build a divergence-free basis, as described in [174] for $k = 1$. When $k = 1$, following the initial work of [42], one can also add P^1-Lagrange basis functions to the space of the discrete pressures as explained in

[P7]. The consistency of the discrete velocity is then improved. Note that Scott-Vogelius finite elements [230, 221, 149] produce velocity approximations that are divergence free.

The code used to get the numerical results can be downloaded on GitHub [G1].

7.5 $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ finite element [P7, W1]

It is well-known that the $\mathbf{P}^1_{nc} \times P^0$ is inaccurate when solving Navier-Stokes equations [145, 180]. In order to improve the accuracy, Bernardi and Hecht propose in [42] to replace $Q_{0,h}$ by the space of P^1-Lagrange pressure (denoted here by $Q_{1,h}$ and defined in Eq. (7.5.1) below) plus local bubble functions, defined in each element as the product of the barycentric coordinates. However, for this discretisation, convergence of the discrete pressure is not guaranteed, as shown in Heib's thesis [175]. In this thesis, a new element is introduced, for which the space of the discrete pressures is the following:

$$Q_h = Q_{0,h} \oplus Q_{1,h}, \text{ where } \left\{ \begin{array}{rcl} Q_{0,h} & := & \mathcal{P}^0_{disc}(\mathcal{T}_h) \cap L^2_{zmv}(\Omega) \\ Q_{1,h} & := & \mathcal{P}^1(\mathcal{T}_h) \cap L^2_{zmv}(\Omega) \end{array} \right. . \tag{7.5.1}$$

The sum is direct: Let $(q_0, q_1) \in Q_{0,h} \times Q_{1,h}$ be a couple such that $q_0 + q_1 = 0$. Then, we have $q_1 = -q_0$ over $\mathcal{T}_h$. Since q_1 is continuous over $\mathcal{T}_h$, it means that $q_1 = -q_0$ are constant. They are equal to zero since $Q_{k,h} \subset L^2_{zmv}(\Omega)$. Let us set:

$$\tilde{Q}_{1,h} = \left\{ q \in Q_h \,|\, \forall \ell \in \mathcal{I}_K, \int_{K_\ell} q = 0 \right\}. \tag{7.5.2}$$

Proposition 7.5.1. *The space $\tilde{Q}_{1,h}$ is such that $Q_h = Q_{0,h} \overset{\perp}{\oplus} \tilde{Q}_{1,h}$, and we have:*

$$\forall q \in Q_h, \quad \|q\|^2_{L^2(\Omega)} = \|\pi_0(q)\|^2_{L^2(\Omega)} + \|q - \pi_0(q)\|^2_{L^2(\Omega)}. \tag{7.5.3}$$

Proof. Let $q \in Q_h$. We set: $q = q_0 + q_1$, where for $k \in \{0, 1\}$, $q_k \in Q_{k,h}(\mathcal{T}_h)$. Notice that: $q - \pi_0(q) = q_1 - \pi_0(q_1) + q_0 - \pi_0(q_0) = q_1 - \pi_0(q_1)$, since $q_0 = \pi_0(q_0)$. Setting $\tilde{q} = q_1 - \pi_0(q_1)$, we have: $(\pi_0(q), \tilde{q})_{L^2(\Omega)} = \sum_{\ell \in \mathcal{I}_K} \pi_0(q)(K_\ell) \int_{K_\ell} \tilde{q} = 0$, so that: $\|q\|^2_{L^2(\Omega)} = \|\pi_0(q) + \tilde{q}\|^2_{L^2(\Omega)} = \|\pi_0(q)\|^2_{L^2(\Omega)} + \|\tilde{q}\|^2_{L^2(\Omega)}$. ⋄

7.5.1 Stability of the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ finite element

In order to prove the inf-sup condition of the finite element $\mathbf{X}_{0,cr} \times Q_h$, we need to construct an explicit right inverse of the divergence operator, from Q_h to $\mathbf{X}_{0,cr}$, cf. Lemma 7.4.1 and from $\tilde{Q}_{1,h}$ to $\mathbf{X}_{0,cr}$, cf. Lemma 7.5.1, under Hypothesis [42, Hypothesis 4.1] given below:

Hypothesis 7.5.1. *We suppose that the triangulation $\mathcal{T}_h$ is such that the boundary $\partial\Omega$ contains at most one edge in dimension $d = 2$ and at most two faces in dimension $d = 3$, of the same element K_ℓ, $\ell \in \mathcal{I}_K$.*

Let us set: $\mathbf{X}^T_{0,cr} := \{\mathbf{v} \in \mathbf{X}_{0,cr} \,|\, \forall f \in \mathcal{I}_F, \underline{\mathbf{v}}_f \cdot \mathbf{n}_f = 0\}$. By construction, $\mathbf{X}^T_{0,cr} \subset \mathbf{V}_{0,cr}$ since:

$$\forall \mathbf{v} \in \mathbf{X}^T_{0,cr}, \forall \ell \in \mathcal{I}_K, \quad \int_{K_\ell} \operatorname{div} \mathbf{v} = - \sum_{f \in \mathcal{I}_{F,\ell}} \int_{F_f} \mathbf{v} \cdot \mathbf{n}_{f,\ell} = \sum_{f \in \mathcal{I}_{F,\ell}} |F_f| \underline{\mathbf{v}}_f \cdot \mathbf{n}_{f,\ell} = 0. \tag{7.5.4}$$

Lemma 7.5.1. *Under Hypothesis 7.5.1, for all $q_1 \in Q_{1,h}$, there exists $\mathbf{v}_{q_1} \in \mathbf{X}^T_{0,cr}$ such that:*

$$(\operatorname{div} \mathbf{v}_{q_1}, \pi_0(q))_{L^2(K_\ell)} = 0 \quad \forall q \in Q_h, \tag{7.5.5}$$

$$(\nabla\, q_1\,,\, \mathbf{v}_{q_1})_{\mathbf{L}^2(\Omega)} \gtrsim \nu^{-1}\, \sigma^{-2\,(d-2)}\, \|q_1 - \pi_0(q_1)\|^2_{L^2(\Omega)}, \tag{7.5.6}$$

$$\|\mathbf{v}_{q_1}\|_h \lesssim \nu^{-1}\, \sigma^2\, \|q_1 - \pi_0(q_1)\|_{L^2(\Omega)}. \tag{7.5.7}$$

Proof. Let us consider $q_1 \in Q_{1,h}$. For all $\ell \in \mathcal{I}_K$, we set $q_{1,\ell} = q_{1|K_\ell}$ and $\hat{q}_{1,\ell} = q_1 \circ T_\ell$. Since $q_1 \in H^1(\Omega)$ we have [16,

Proposition 2.2.10]: for all $f \in \mathcal{I}_F^i$ such that $F_f = K_\ell \cap K_{\ell'}$, $\nabla\, q_{1,\ell} \times \mathbf{n}_f = \nabla\, q_{1,\ell'} \times \mathbf{n}_f$.
We can then construct $\mathbf{v}_{q_1} \in \mathbf{X}_{0,h}^T$ such that[3]:

$$\forall f \in \mathcal{I}_F^i \quad \mathbf{v}_{q_1}(M_f) = \nu^{-1}\,(h_f)^2\,\mathbf{n}_f \times (\nabla\, q_1 \times \mathbf{n}_f). \tag{7.5.8}$$

Since $\mathbf{v}_{q_1} \in \mathbf{X}_{0,h}^T$, Equation (7.5.5) holds.

Let us prove (7.5.6). Noticing that $\forall \ell \in \mathcal{I}_K$, $\forall f \in \mathcal{I}_{F,\ell}$, $\displaystyle\int_{K_\ell} \psi_f = \frac{|K_\ell|}{d+1}$, we have:

$$\begin{aligned}
(\nabla\, q_1, \mathbf{v}_{q_1})_{\mathbf{L}^2(K_\ell)} &= (d+1)^{-1}|K_\ell| \sum_{f \in \mathcal{I}_{F,\ell}^i} \nabla\, q_{1,\ell} \cdot \mathbf{v}_{q_1}(M_f),\\
&= \nu^{-1}(d+1)^{-1}|K_\ell| \sum_{f \in \mathcal{I}_{F,\ell}^i} (h_f)^2\ \nabla\, q_{1,\ell} \cdot \mathbf{n}_f \times (\nabla\, q_{1,\ell} \times \mathbf{n}_f),\\
&= \nu^{-1}(d+1)^{-1}|K_\ell| \sum_{f \in \mathcal{I}_{F,\ell}^i} (h_f)^2\ |\nabla\, q_{1,\ell} \times \mathbf{n}_f|^2\,.
\end{aligned} \tag{7.5.9}$$

Using Equations (7.3.4) and (7.3.10), we have: $\nabla\, q_{1,\ell} \times \mathbf{n}_f = |J_\ell|\,|J_{f,\ell}|^{-1}\,(\mathbb{B}_\ell{}^{-1})^T\,\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times (\mathbb{B}_\ell{}^{-1})^T\,\mathbf{n}_{\hat{F}_{f,\ell}}$. By a direct computation, we show that: $(\mathbb{B}_\ell{}^{-1})^T\,\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times (\mathbb{B}_\ell{}^{-1})^T\,\mathbf{n}_{\hat{F}_{f,\ell}} = (J_\ell)^{-1}(\mathbb{B}_\ell)^{d-2}\ \nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times \mathbf{n}_{\hat{F}_{f,\ell}}$. Hence:

$$|\,\nabla\, q_{1,\ell} \times \mathbf{n}_f| = |J_{f,\ell}|^{-1}\ \left|\,(\mathbb{B}_\ell)^{d-2}\ \nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times \mathbf{n}_{\hat{F}_{f,\ell}}\right|. \tag{7.5.10}$$

Notice that we have:

$$\left|(\mathbb{B}_\ell)^{d-2}\ \left(\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times \mathbf{n}_{\hat{F}_{f,\ell}}\right)\right| \geq \left\|\mathbb{B}_\ell{}^{-1}\right\|^{-(d-2)}\ \left|\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times \mathbf{n}_{\hat{F}_{f,\ell}}\right| \gtrsim (\rho_\ell)^{d-2}\ \left|\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times \mathbf{n}_{\hat{F}_{f,\ell}}\right|. \tag{7.5.11}$$

Using Equation (7.5.11) in (7.5.10) times h_f, squared, we get: $(h_f)^2\,|\nabla\, q_{1,\ell} \times \mathbf{n}_f|^2 \gtrsim (|J_{f,\ell}|^{-1}\,h_f)^2\,(\rho_\ell)^{2\,(d-2)}\ \left|\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times \mathbf{n}_{\hat{F}_{f,\ell}}\right|^2$.
Notice that $(|J_{f,\ell}|^{-1}\,h_f)^2 \gtrsim (h_f)^{-2\,(d-2)} \geq (h_\ell)^{-2\,(d-2)}$, hence: $(|J_{f,\ell}|^{-1}\,h_f)^2\,(\rho_\ell)^{2\,(d-2)} \gtrsim \left(\frac{h_\ell}{\rho_\ell}\right)^{-2\,(d-2)} \geq (\sigma_\ell)^{-2\,(d-2)}$.
It follows that:

$$(h_f)^2\ |\nabla\, q_{1,\ell} \times \mathbf{n}_f|^2 \gtrsim (\sigma_\ell)^{-2\,(d-2)}\ \left|\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times \mathbf{n}_{\hat{F}_f}\right|^2. \tag{7.5.12}$$

According to Hypothesis 7.5.1, we have for all $\ell \in \mathcal{I}_K$:

$$\sum_{f \in \mathcal{I}_{F,\ell}^i} \left|\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell} \times \mathbf{n}_{\hat{F}_f}\right|^2 \gtrsim |\,\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell}|^2. \tag{7.5.13}$$

We can now bound the left-hand term in (7.5.9) using Equations (7.5.12) and (7.5.13):

$$(\nabla\, q_1, \mathbf{v}_{q_1})_{\mathbf{L}^2(K_\ell)} \gtrsim \nu^{-1}\,(\sigma_\ell)^{-2\,(d-2)}|K_\ell|\,|\,\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell}|^2. \tag{7.5.14}$$

Using inequality (7.3.7) in the reference element and the change of variable (7.3.3), we have:

$$\begin{aligned}
|K_\ell|\,|\,\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell}|^2 &= |K_\ell|\,|\hat{K}|^{-1}\,\|\,\nabla_{\hat{\mathbf{x}}}\,\hat{q}_{1,\ell}\|^2_{\mathbf{L}^2(\hat{K})},\\
&\gtrsim |K_\ell|\,\|\hat{q}_{1,\ell} - \underline{q}_{1,\ell}\|^2_{L^2(\hat{K})},\\
&\gtrsim \|q_{1,\ell} - \underline{q}_{1,\ell}\|^2_{L^2(K_\ell)} \text{ where } \underline{q}_{1,\ell} = |K_\ell|^{-1}\,\textstyle\int_{K_\ell} q_1.
\end{aligned} \tag{7.5.15}$$

Thus, using (7.5.15) in (7.5.14), and summing over $\ell \in \mathcal{I}_K$, we obtain (7.5.6).

Let us prove (7.5.7). For all $\ell \in \mathcal{I}_K$, $\boldsymbol{\nabla}\,\mathbf{v}_{q_1|K_\ell} = \displaystyle\sum_{f \in \mathcal{I}_{F,\ell}^i} \mathbf{v}_{q_1}(M_f) \otimes \nabla\,\psi_{f|K_\ell}$. We deduce that:

$$\begin{aligned}
\|\,\boldsymbol{\nabla}\,\mathbf{v}_{q_1}\|^2_{\mathbb{L}^2(K_\ell)} &\lesssim |K_\ell| \sum_{f \in \mathcal{I}_{F,\ell}^i} |\mathbf{v}_{q_1}(M_f)|^2\,|\,\nabla\,\psi_{f|K_\ell}|^2,\\
&\lesssim \nu^{-2}\,|K_\ell| \sum_{f \in \mathcal{I}_{F,\ell}^i} (h_f)^4\ |\mathbf{n}_f \times \nabla\, q_1 \times \mathbf{n}_f|^2\ (|K_\ell|^{-2}\,|F_f|^2) && \text{since } |\,\nabla\,\psi_{f|K_\ell}|^2 = |K_\ell|^{-2}\,|F_f|^2,\\
&\lesssim \nu^{-2}\,|K_\ell|\,(h_\ell)^4\,(\rho_\ell)^{-2} \sum_{f \in \mathcal{I}_{F,\ell}^i} |\mathbf{n}_f \times \nabla\, q_1 \times \mathbf{n}_f|^2 && \text{using (7.3.11)},\\
&\lesssim \nu^{-2}\,(\sigma_\ell\,h_\ell)^2\,|K_\ell|\,|\,\nabla\, q_1|^2 = \nu^{-2}\,(\sigma_\ell\,h_\ell)^2\,\|\,\nabla\, q_1\|^2_{\mathbf{L}^2(K_\ell)} && \text{using Hypothesis 7.5.1},\\
&\lesssim \nu^{-2}\,(\sigma_\ell)^4\,\|q_{1,\ell} - \underline{q}_{1,\ell}\|^2_{L^2(K_\ell)} && \text{using (7.3.14)}.
\end{aligned}$$

$\diamond$

[3] In $2D$, $\nabla\, q_1 \times \mathbf{n}_f$ is a scalar, hence we make the following abuse of notation: $\mathbf{n}_f \times (\nabla\, q_1 \times \mathbf{n}_f) = (\nabla\, q_1 \times \mathbf{n}_f)\tau_f$ where $(\tau_f, \mathbf{n}_f)$ is an orthogonal basis.

Let b_h be the bilinear form such that:

$$b_h : \left\{ \begin{array}{rcl} \mathbf{X}_h \times Q_h & \mapsto & \mathbb{R} \\ (\mathbf{v}, q) & \to & -\left(\operatorname{div}_h \mathbf{v}, q_0 \right)_{L^2(\Omega)} + (\mathbf{v}, \nabla\, q_1)_{\mathbf{L}^2(\Omega)}, \end{array} \right. \tag{7.5.16}$$

with $q = q_0 + q_1$ such that $(q_0, q_1) \in Q_{0,h} \times Q_{1,h}$. Under Hypothesis 7.5.1, the bilinear form $b_h(\cdot,\cdot)$ is continuous and is satisfies a discrete inf-sup condition, as established in

Theorem 7.5.1 (`Stability`). *The following continuity property holds:*

$$\forall \mathbf{v} \in \mathbf{X}_{0,h},\, \forall q \in Q_h, \quad |b_h(\mathbf{v}, q)| \lesssim \sigma\, \|\mathbf{v}\|_h\, \|q\|_{L^2(\Omega)}. \tag{7.5.17}$$

Under Hypothesis 7.5.1, the following inf-sup condition applies

$$\forall q \in Q_h,\, \exists \mathbf{v} \in \mathbf{X}_{0,h}, \quad b_h(\mathbf{v}, q) \gtrsim C_{\mathrm{div}}^{-2}\, \sigma^{-2\,d}\, \|\mathbf{v}\|_h\, \|q\|_{L^2(\Omega)}. \tag{7.5.18}$$

Proof. Let us prove (7.5.17). Consider $(\mathbf{v}, q) \in \mathbf{X}_{0,h} \times Q_h$, where $q = q_0 + q_1$, $(q_0, q_1) \in Q_{0,h} \times Q_{1,h}$. Integrating by parts the last term of (7.5.16), we obtain that:

$$b_h(\mathbf{v}, q) = -(\operatorname{div}_h \mathbf{v}, \pi_0(q))_{L^2(\Omega)} + \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\mathbf{v} \cdot \mathbf{n}_{f,\ell}\; q_1)_{L^2(F_f)}. \tag{7.5.19}$$

On the one hand, using Cauchy-Schwarz inequality, we have:

$$|(\operatorname{div}_h \mathbf{v}, \pi_0(q))_{L^2(\Omega)}| \leq \sqrt{d}\, \|\mathbf{v}\|_h\, \|\pi_0(q)\|_{L^2(\Omega)} \tag{7.5.20}$$

On the other hand, using inequality (7.4.3), it holds:

$$\left| \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\mathbf{v} \cdot \mathbf{n}_{f,\ell}\; q_1)_{L^2(F_f)} \right| \lesssim \sigma\, \|\mathbf{v}\|_h\, \|q_1 - \pi_0(q_1)\|_{L^2(\Omega)} = \sigma\, \|\mathbf{v}\|_h\, \|q - \pi_0(q)\|_{L^2(\Omega)}. \tag{7.5.21}$$

We obtain (7.5.17) summing (7.5.20) and (7.5.21), and using (7.5.3).

Let us prove (7.5.18). Consider $q \neq 0$. Let $\mathbf{v}_{\pi_0(q)}$ be built like $\mathbf{v}_{h,q_0}$ in Lemma 7.4.1 with $q_0 = \pi_0(q)$ here and $\mathbf{v}_{q_1}$ be built like in Lemma 7.5.1. We have:

$$\begin{aligned} -\left(\operatorname{div}_h \mathbf{v}_{\pi_0(q)}, \pi_0(q) \right)_{L^2(\Omega)} &= \nu^{-1} \|\pi_0(q)\|^2_{L^2(\Omega)}, && (7.5.22) \\ \|\mathbf{v}_{\pi_0(q)}\|_h &\leq \nu^{-1}\, C_{\mathrm{div}}\, \|\pi_0(q)\|_{L^2(\Omega)}, && (7.5.23) \\ (\operatorname{div}_h \mathbf{v}_{q_1}, \pi_0(q))_{L^2(\Omega)} &= 0, && (7.5.24) \\ (\nabla\, q_1\, ,\, \mathbf{v}_{q_1})_{\mathbf{L}^2(\Omega)} &\gtrsim \nu^{-1}\, \sigma^{-2\,(d-2)}\, \|q - \pi_0(q)\|^2_{L^2(\Omega)}, && (7.5.25) \\ \|\mathbf{v}_{q_1}\|_h &\lesssim \nu^{-1}\, \sigma^2\, \|q - \pi_0(q)\|_{L^2(\Omega)}. && (7.5.26) \end{aligned}$$

Using equality (7.5.19) with $\mathbf{v} = \mathbf{v}_{\pi_0(q)}$, we have, due to (7.5.22):

$$b_h(\mathbf{v}_{\pi_0(q)}, q) = \nu^{-1} \|\pi_0(q)\|^2_{L^2(\Omega)} + \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\mathbf{v}_{\pi_0(q)} \cdot \mathbf{n}_{f,\ell}\; q_1)_{L^2(F_f)}.$$

Using inequality (7.5.21) with $\mathbf{v} = \mathbf{v}_{\pi_0(q)}$, and applying (7.5.23), it holds:

$$b_h(\mathbf{v}_{\pi_0(q)}, q) \gtrsim \nu^{-1} \left(\|\pi_0(q)\|^2_{L^2(\Omega)} - C^{\sigma}_{\mathrm{div}}\, \|\pi_0(q)\|_{L^2(\Omega)}\, \|q - \pi_0(q)\|_{L^2(\Omega)} \right) \text{ with } C^{\sigma}_{\mathrm{div}} = C_{\mathrm{div}}\, \sigma.$$

Using Young inequality, we have for all $\varepsilon > 0$:

$$b_h(\mathbf{v}_{\pi_0(q)}, q) \geq \nu^{-1} \left(1 - \frac{\varepsilon}{2} C^{\sigma}_{\mathrm{div}} \right) \|\pi_0(q)\|^2_{L^2(\Omega)} - \nu^{-1}\, \frac{\varepsilon^{-1}}{2} C^{\sigma}_{\mathrm{div}}\, \|q - \pi_0(q)\|^2_{L^2(\Omega)}. \tag{7.5.27}$$

Now, using Equations (7.5.24) and (7.5.25), we have:

$$b_h(\mathbf{v}_{q_1}, q) = (\nabla\, q_1, \mathbf{v}_{q_1})_{\mathbf{L}^2(\Omega)} \gtrsim \nu^{-1} \sigma^{-2(d-2)}\, \|q - \pi_0(q)\|^2_{L^2(\Omega)}. \tag{7.5.28}$$

Let us set $\mathbf{v}^\star = \mu \mathbf{v}_{\pi_0(q)} + \mathbf{v}_{q_1}$, where $\mu > 0$ is a constant. We have: $b_h(\mathbf{v}^\star, q) = \mu\, b_h(\mathbf{v}_{\pi_0(q)}, q) + b_h(\mathbf{v}_{q_1}, q)$. Summing Equation

(7.5.27) times μ with Equation (7.5.28), we obtain:

$$b_h(\mathbf{v}^\star, q) \gtrsim \mu\,\nu^{-1}\left(1-\frac{\varepsilon}{2}\,C_{\rm div}^{\sigma}\right)\|\pi_0(q)\|_{L^2(\Omega)}^2+\nu^{-1}\left(\sigma^{-2\,(d-2)}-\frac{\mu\,\varepsilon^{-1}}{2}\,C_{\rm div}^{\sigma}\right)\|q-\pi_0(q)\|_{L^2(\Omega)}^2.$$

We need to enforce that $1-\frac{\varepsilon}{2}C_{\rm div}^{\sigma}>0$ and $\sigma^{-2\,(d-2)}-\frac{\mu\,\varepsilon^{-1}}{2}\,C_{\rm div}^{\sigma}>0$, which amounts to requiring

$$\varepsilon<2\,(C_{\rm div}^{\sigma})^{-1}\text{ and }\mu<2\,(C_{\rm div}^{\sigma})^{-1}\,\sigma^{-2(d-2)}\varepsilon.$$

Setting for instance $\varepsilon=(C_{\rm div}^{\sigma})^{-1}$ and $\mu=(C_{\rm div}^{\sigma})^{-1}\,\varepsilon\,\sigma^{-2\,(d-2)}$, we obtain:

$$b_h(\mathbf{v}^\star, q)\gtrsim\frac{1}{2}\left(\mu\,\nu^{-1}\,\|\pi_0(q)\|_{L^2(\Omega)}^2+\nu^{-1}\,\sigma^{-2\,(d-2)}\|q-\pi_0(q)\|_{L^2(\Omega)}^2\right).$$

Since $C_{\rm div}\geq 1$ and $\sigma\geq 1$, we have: $(C_{\rm div}^{\sigma})^{-2}\leq 1$ and $\mu=(C_{\rm div}^{\sigma})^{-2}\,\sigma^{-2\,(d-2)}\leq\sigma^{-2\,(d-2)}$. Hence:

$$b_h(\mathbf{v}^\star, q)\quad\gtrsim\quad\frac{1}{2}\nu^{-1}\,(C_{\rm div})^{-2}\,\sigma^{-2\,(d-1)}\,\left(\|\pi_0(q)\|_{L^2(\Omega)}^2+\|q-\pi_0(q)\|_{L^2(\Omega)}^2\right).\tag{7.5.29}$$

We will now bound $\|\mathbf{v}^\star\|_h$ by $\|q\|_{L^2(\Omega)}$. Using Equations (7.5.23) and (7.5.26), we obtain:

$$\begin{aligned}\|\mathbf{v}^\star\|_h^2&\leq 2\left(\mu^2\,\|\mathbf{v}_{\pi_0(q)}\|_h^2+\|\mathbf{v}_{q_1}\|_h^2\right),\\&\leq 2\,\nu^{-2}\left(\sigma^{-4\,(d-1)}\,(C_{\rm div})^{-2}\,\|\pi_0(q)\|_{L^2(\Omega)}^2+\sigma^4\,\|q-\pi_0(q)\|_{L^2(\Omega)}^2\right),\\&\leq 2\,\nu^{-2}\,\sigma^4\,\|q\|_{L^2(\Omega)}^2.\end{aligned}$$

Hence: $\nu^{-1}\|q\|_{L^2(\Omega)}\geq\frac{1}{\sqrt{2}}\,\sigma^{-2}\,\|\mathbf{v}^\star\|_h$. Using this bound in (7.5.29), we obtain (7.5.18). ◇

Remark 7.5.1. *We exhibit that the bound of the discrete inf-sup condition depends on $C_{\rm div}$ and σ. We could further optimize the choice of ε and μ in the proof above to obtain a sharper bound.*

Let $\mathbf{V}_{cr}\subset\mathbf{V}_{0,cr}$ be the kernel of the bilinear form $b_h(\cdot,\cdot)$:

$$\mathbf{V}_{cr}:=\{\mathbf{v}\in\mathbf{X}_{0,cr}\,|\,\forall q\in Q_h,\,b_h(\mathbf{v},q)=0\}.\tag{7.5.30}$$

The $\mathbf{P}_{nc}^1\times(P^0+P^1)$ finite element method allows to improve accuracy compared to the $\mathbf{P}_{nc}^1\times P^0$ finite element method:

Proposition 7.5.2. *Let $\mathbf{v}\in\mathbf{V}_{cr}$. Then for all $q_1\in Q_{1,h}$: $\displaystyle\sum_{f\in\mathcal{I}_F}(q_1,[\mathbf{v}]_f\cdot\mathbf{n}_f)_{L^2(F_f)}=0$.*

Proof. Let $(\mathbf{v},q_1)\in\mathbf{V}_{cr}\times Q_{1,h}$. We have $(\mathrm{div}_h\,\mathbf{v},\pi_0(q_1))_{L^2(\Omega)}+(\nabla\,q_1,\mathbf{v})_{\mathbf{L}^2(\Omega)}=0$. Integrating by parts on all the elements, we get that:

$$0=(\nabla\,q_1,\mathbf{v})_{\mathbf{L}^2(\Omega)}+(\mathrm{div}_h\,\mathbf{v},\pi_0(q_1))_{L^2(\Omega)}+\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}(q_1,\mathbf{v}\cdot\mathbf{n}_{f,\ell})_{L^2(F_f}=\sum_{f\in\mathcal{I}_F}(q_1,[\mathbf{v}]_f\cdot\mathbf{n}_f)_{L^2(F_f)}.$$

◇

We proved the stability of the $\mathbf{P}_{nc}^1\times(P^0+P^1)$ finite element, cf. Theorem 7.5.1 and we showed that the accuracy is improved with respect to the $\mathbf{P}_{nc}^1\times P^0$ finite element method, cf. Proposition 7.5.2. Hence, we have the tools to prove that the discretization of Problem (7.2.1) with the $\mathbf{P}_{nc}^1\times(P^0+P^1)$ finite element is well posed.

7.5.2 Well-posedness

Let us set $\mathcal{X}_h:=\mathbf{X}_{0,h}\times Q_h$. We endow $\mathcal{X}_h$ with the norm: $\|(\mathbf{v},q)\|_{\mathcal{X}_h}:=(\|\mathbf{v}\|_h^2+\nu^{-2}\,\|q\|_{L^2(\Omega)}^2)^{1/2}$, so that it is a Hilbert space. Consider the symmetric and continuous bilinear form $a_h:\mathcal{X}_h\times\mathcal{X}_h\to\mathbb{R}$ such

that for all $(\,(\mathbf{u}_h', p_h'), (\mathbf{v}_h, q_h)\,) \in \mathcal{X}_h \times \mathcal{X}_h$:

$$a_h(\,(\mathbf{u}_h', p_h'), (\mathbf{v}_h, q_h)\,) = \nu(\mathbf{u}_h', \mathbf{v}_h)_h + b_h(\mathbf{v}_h, p_h') + b_h(\mathbf{u}_h', q_h). \tag{7.5.31}$$

The discrete variational formulation of Problem (7.2.1) with the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ finite element reads:

$$\text{Find } (\mathbf{u}_h, p_h) \in \mathcal{X}_h \text{ such that } \forall (\mathbf{v}_h, q_h) \in \mathcal{X}_h, a_h\,(\,(\mathbf{u}_h, p_h), (\mathbf{v}_h, q_h)\,) = \ell_{\mathbf{f}}(\mathbf{v}_h), \tag{7.5.32}$$

The main consequence of the inf-sup condition (7.5.18) is that Problem (7.5.32) is well posed [50]. We have the following result.

Theorem 7.5.2 (`Stability of the solution`)**.** *If Hypothesis 7.5.1 holds, for any data* $\mathbf{f}$*, Problem* (7.5.32) *has a unique solution* $(\mathbf{u}_h, p_h) \in \mathcal{X}_h$ *such that:*

$$\|\mathbf{u}_h\|_h \lesssim \nu^{-1}\,\sigma\,C_{\mathbf{f}}, \tag{7.5.33}$$

$$\|p_h - \pi_0(p_h)\|_{L^2(\Omega)} \lesssim 2\,\sigma^{2d-1}\,C_{\mathbf{f}}, \tag{7.5.34}$$

$$\|\pi_0(p_h)\|_{L^2(\Omega)} \lesssim 2\,\sigma^{2d}\,C_{\mathrm{div}}\,C_{\mathbf{f}}. \tag{7.5.35}$$

where $C_{\mathbf{f}} = C_\Omega\,h_\Omega\,\|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}$ *if* $\mathbf{f} \in \mathbf{L}^2(\Omega)$*, or* $C_{\mathbf{f}} = C_{\mathcal{J}_h}\,\|\mathbf{f}\|_{\mathbf{H}^{-1}(\Omega)}$ *if* $\mathbf{f} \notin \mathbf{L}^2(\Omega)$*.*

Proof. We write the proof for $\mathbf{f} \in \mathbf{L}^2(\Omega)$. It is very similar when $\mathbf{f} \in \mathbf{H}^{-1}(\Omega)$, using inequality (7.4.7).

Let $(\mathbf{u}_h, p_h)$ be the unique solution to Problem (7.5.32), with $p_h = p_{0,h} + p_{1,h}$, $(p_{0,h}, p_{1,h}) \in Q_{0,h} \times Q_{1,h}$.

Let us prove (7.5.33). Choosing $\mathbf{v}_h = 0$ in (7.5.32), we have: $\mathbf{u}_h \in \mathbf{V}_{cr}$. Choosing $(\mathbf{v}_h, q_h) = (\mathbf{u}_h, 0)$ in (7.5.32), we then obtain: $\nu\|\mathbf{u}_h\|_h^2 = (\mathbf{f}, \mathbf{u}_h)_{\mathbf{L}^2(\Omega)}$. Using Cauchy-Schwarz inequality and the discrete Poincaré-Steklov inequality (7.4.4), we get inequality (7.5.33).

Let us prove (7.5.34). Choosing $(\mathbf{v}_h, q_h) = (\mathbf{v}_{p_{1,h}}, 0)$ in (7.5.32), where $\mathbf{v}_{p_{1,h}}$ is constructed like $\mathbf{v}_{q_1}$ in Lemma 7.5.1 with $q_1 = p_{1,h}$ here, and reshuffling the terms, we obtain:

$$(\nabla\,p_{1,h}, \mathbf{v}_{p_{1,h}})_{\mathbf{L}^2(\Omega)} = (\mathbf{f}, \mathbf{v}_{p_{1,h}})_{\mathbf{L}^2(\Omega)} - \nu(\mathbf{u}_h, \mathbf{v}_{p_{1,h}})_h.$$

Using the triangular inequality and Cauchy-Schwarz, we get:

$$\left|(\mathbf{v}_{p_{1,h}}, \nabla\,p_{1,h})_{\mathbf{L}^2(\Omega)}\right| \le \|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\,\|\mathbf{v}_{p_{1,h}}\|_{\mathbf{L}^2(\Omega)} + \nu\|\mathbf{u}_h\|_h\;\|\mathbf{v}_{p_{1,h}}\|_h.$$

Now, using the discrete Poincaré-Steklov inequality (7.4.4) and inequality (7.5.33), we have:

$$\left|(\mathbf{v}_{p_{1,h}}, \nabla\,p_{1,h})_{\mathbf{L}^2(\Omega)}\right| \le 2\,\sigma\,C_\Omega\,h_\Omega\,\|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\,\|\mathbf{v}_{p_{1,h}}\|_h.$$

Using inequality (7.5.7), we deduce that:

$$\left|(\mathbf{v}_{p_{1,h}}, \nabla\,p_{1,h})_{\mathbf{L}^2(\Omega)}\right| \lesssim 2\,\nu^{-1}\,\sigma^3\,C_\Omega\,h_\Omega\,\|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\,\|p_{1,h} - \pi_0(p_{1,h})\|_{L^2(\Omega)}.$$

Hence, using inequality (7.5.6), we get:

$$\nu^{-1}\,\sigma^{-2(d-2)}\,\|p_{1,h} - \pi_0(p_{1,h})\|_{L^2(\Omega)}^2 \lesssim 2\,\nu^{-1}\,\sigma^3\,C_\Omega\,h_\Omega\,\|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\,\|p_{1,h} - \pi_0(p_{1,h})\|_{L^2(\Omega)}.$$

We conclude that:

$$\|p_h - \pi_0(p_h)\|_{L^2(\Omega)} = \|p_{1,h} - \pi_0(p_{1,h})\|_{L^2(\Omega)} \lesssim 2\,\sigma^{2\,d-1}\,C_\Omega\,h_\Omega\,\|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}.$$

Let us prove (7.5.35). Consider $(\mathbf{v}_h, q_h) = (\mathbf{v}_{\pi_0(p_h)}, 0)$ in (7.5.32), where $\mathbf{v}_{\pi_0(p_h)}$ is built like $\mathbf{v}_{q_{0,h}}$ in Lemma 7.4.1 with $q_{0,h} = \pi_0(p_h)$. Using Equation (7.5.19) with $q = p_h$, Equation (7.4.37) and reshuffling the terms, it holds:

$$\nu^{-1}\|\pi_0(q)\|_{L^2(\Omega)}^2 = (\mathbf{f}, \mathbf{v}_{\pi_0(p_h)})_{\mathbf{L}^2(\Omega)} - \nu(\mathbf{u}_h, \mathbf{v}_{\pi_0(p_h)})_h - \sum_{\ell \in \mathcal{I}_K}\sum_{f \in \mathcal{I}_{F,\ell}} (\mathbf{v}_{\pi_0(p_h)} \cdot \mathbf{n}_{f,\ell}, p_{1,h})_{L^2(F_f)}.$$

Using Cauchy-Schwarz and inequality (7.4.3), we have:

$$\nu^{-1}\|\pi_0(q)\|_{L^2(\Omega)}^2 \le \|\mathbf{f}\|_{\mathbf{L}^2(\Omega)}\,\|\mathbf{v}_{\pi_0(p_h)}\|_{\mathbf{L}^2(\Omega)} + \nu\|\mathbf{u}_h\|_h\,\|\mathbf{v}_{\pi_0(p_h)}\|_h + \sigma\,\|\mathbf{v}_{\pi_0(p_h)}\|_h\,\|p_{1,h} - \pi_0(p_{1,h})\|_{L^2(\Omega)}.$$

Using the discrete Poincaré-Steklov inequality (7.4.4) and inequality (7.4.38), it holds:

$$\|\pi_0(q)\|_{L^2(\Omega)}^2 \le C_{\mathrm{div}}\,\left(\sigma C_\Omega\,h_\Omega\,\|\mathbf{f}\|_{\mathbf{L}^2(\Omega)} + \nu\|\mathbf{u}_h\|_h\; + \sigma\,\|p_{1,h} - \pi_0(p_{1,h})\|_{L^2(\Omega)}\right)\,\|\pi_0(q)\|_{L^2(\Omega)}.$$

Using inequalities (7.5.33) and (7.5.34), we deduce (7.5.35). ◇

In order to compute the error estimates, we need to exhibit a Fortin operator.

7.5.3 Fortin operator

In this section, we construct a Fortin operator, inspired by [42, Proposition 6.3]. Let us split $\mathbf{X}_{0,cr}^T$ this way: $\mathbf{X}_{0,cr}^T = \mathbf{V}_{cr}^T \oplus \mathbf{W}_{cr}^T$, where $\mathbf{V}_{cr}^T = \mathbf{V}_{cr} \cap \mathbf{X}_{0,cr}^T$ and $\mathbf{W}_{cr}^T$ is the orthogonal of $\mathbf{V}_{cr}^T$ in $\mathbf{X}_{0,cr}^T$. We make the abuse of using the same notation for $b_h(\cdot,\cdot)$ and its extension to $(\mathbf{X}_{0,cr} + \mathbf{H}_0^1(\Omega)) \times Q_h$.

Proposition 7.5.3 (`Fortin operator`). *Consider the projection operator $\Pi_h : \mathbf{H}_0^1(\Omega) \to \mathbf{X}_{0,cr}$ such that: for all $\mathbf{w} \in \mathbf{H}_0^1(\Omega)$, $\Pi_h \mathbf{w} = \Pi_{cr}\mathbf{w} + \mathbf{w}_T$, where $\mathbf{w}_T \in \mathbf{W}_{cr}^T$ satisfies:*

$$\forall q_1 \in Q_{1,h}, \quad (\mathbf{w}_T, \nabla\, q_1)_{\mathbf{L}^2(\Omega)} = (\mathbf{w} - \Pi_{cr}\mathbf{w}, \nabla\, q_1)_{\mathbf{L}^2(\Omega)}. \tag{7.5.36}$$

Then, the projection operator Π_h is such that:

$$\forall \mathbf{w} \in \mathbf{H}_0^1(\Omega) \qquad b_h(\Pi_h \mathbf{w}, q_h) = b_h(\mathbf{w}, q_h), \quad \forall q \in Q_h. \tag{7.5.37}$$

$$\forall \mathbf{w} \in \mathbf{H}^{s+1}(\Omega) \cap \mathbf{H}_0^1(\Omega) \quad \|\mathbf{w} - \Pi_h \mathbf{w}\|_h \lesssim \sigma^{2d}\, h^s |\mathbf{w}|_{\mathbf{H}^{s+1}(\Omega)}, \quad s \in [0,1]. \tag{7.5.38}$$

Proof. Let $\mathbf{w} \in \mathbf{H}_0^1(\Omega)$. Let $q = (q_0, q_1) \in Q_{0,h} \times Q_{1,h}$. It holds:

$$-(\mathrm{div}_h\, \Pi_h \mathbf{w}, q_0)_{L^2(\Omega)} = -(\mathrm{div}_h\, \Pi_{cr}\mathbf{w}, q_0)_{L^2(\Omega)} = -(\mathrm{div}\, \mathbf{w}, q_0)_{L^2(\Omega)}.$$

Due to Lemma 7.5.1, $\mathbf{w}_T$ exists and is unique. By construction, Equation (7.5.36) implies (7.5.37). Let $\mathbf{w} \in \mathbf{H}^{1+s}(\Omega) \cap \mathbf{H}_0^1(\Omega)$. Integrating by parts, we have, for all $q_1 \in Q_{1,h}$:

$$(\mathbf{w}_T, \nabla\, q_1)_{\mathbf{L}^2(\Omega)} = -\sum_{\ell \in \mathcal{I}_K} (\mathrm{div}(\mathbf{w} - \Pi_{cr}\mathbf{w}), q_1 - \pi_0(q_1))_{L^2(K_\ell)} + \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\,(\mathbf{w} - \Pi_{cr}\mathbf{w}) \cdot \mathbf{n}_{f,\ell}, q_1 - \pi_0(q_1))_{L^2(F_f)}.$$

Hence, using (7.4.3) and the discrete Cauchy-Schwarz inequality, we deduce that:

$$|(\mathbf{w}_T, \nabla\, q_1)_{\mathbf{L}^2(\Omega)}| \lesssim (\sigma + \sqrt{d})\, \|\mathbf{w} - \Pi_{cr}\mathbf{w}\|_{\mathbf{L}^2(\Omega)}\, \|q_1 - \pi_0(q_1)\|_{L^2(\Omega)}.$$

Using (7.4.34) and the discrete Cauchy-Schwarz inequality, we obtain that:

$$|(\mathbf{w}_T, \nabla\, q_1)_{\mathbf{L}^2(\Omega)}| \lesssim \sigma^2\, h^s\, |\mathbf{w}|_{\mathbf{H}^{1+s}(\Omega)}\, \|q_1 - \pi_0(q_1)\|_{L^2(K_\ell)}. \tag{7.5.39}$$

According to [161, Chapter I, Lemma 4.1 (iii)], thanks to Lemma 7.5.1, we have the following estimate:

$$\|\mathbf{w}_T\|_h \le \sigma^{2(d-1)} \sup_{q_1 \in Q_{1,h} \setminus \{0\}} \frac{|(\mathbf{w}_T, \nabla\, q_1)_{\mathbf{L}^2(\Omega)}|}{\|q_1 - \pi_0(q_1)\|_{L^2(\Omega)}}. \tag{7.5.40}$$

Using (7.5.39) in (7.5.40), we deduce that:

$$\|\mathbf{w}_T\|_h \lesssim \sigma^{2d}\, h^s\, |\mathbf{w}|_{\mathbf{H}^{1+s}(\Omega)}. \tag{7.5.41}$$

We obtain (7.5.38) from (7.4.34) and (7.5.41), since $\|\mathbf{w} - \Pi_h \mathbf{w}\|_h \le \|\mathbf{w} - \Pi_{cr}\mathbf{w}\|_h + \|\mathbf{w}_T\|_h$. ◇

We exhibited a Fortin operator. Let us now study the convergence of the discretization of Problem (7.2.1) with the $\mathbf{P}_{nc}^1 \times (P^0 + P^1)$ finite element.

7.5.4 Classical error estimates

Definition 7.5.1. *Let π_1 be the L^2-projection to $Q_{1,h}$. Let $q \in Q_{1,h}$. One can set $\pi_1(q) = \tilde{\pi}_1(q) - |\Omega|^{-1} \int_\Omega \tilde{\pi}_1(q)$, where $\tilde{\pi}_1$ is the L^2-projection to $\mathcal{P}^1(\mathcal{T}_h)$. It holds [146, Lemma 1.131][4]:*

$$|\pi_1(q)|_{H^1(\Omega)} \lesssim \sigma |q|_{H^1(\Omega)}. \tag{7.5.42}$$

[4] cf. the proof of [248, Lemma 5.1.1], which uses Scott-Zhang interpolation [231]. We can exhibit σ from [231, Theorem 1].

Using the triangular inequality and (7.5.42), it holds:

$$|\pi_1(q) - q|_{H^1(\Omega)} \leq |q|_{H^1(\Omega)} + |\pi_1(q)|_{H^1(\Omega)} \lesssim (\sigma + 1)\, |q|_{H^1(\Omega)}. \tag{7.5.43}$$

Let $q \in L^2_{zmv}(\Omega) \cap H^1(\Omega)$. Using inequality (7.3.18), and (7.5.43), it holds:

$$\|(\pi_1(q) - q) - \pi_0(\pi_1(q) - q)\|_{L^2(\Omega)} \leq \pi^{-1}\, h\, (\sigma + 1)\, |q|_{H^1(\Omega)}. \tag{7.5.44}$$

We have the tools to prove the convergence of the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ discretization. In what follows, we set: $\|(\mathbf{u}, p)\|_\nu = |\mathbf{u}|_{\mathbf{H}^2(\Omega)} + \nu^{-1}\, |p|_{H^1(\Omega)}$.

Theorem 7.5.3 (`Convergence of the discrete velocity`). *Suppose that Hypothesis 7.5.1 holds and that the solution to Problem* (7.2.1) *is such that* $(\mathbf{u}, p) \in \big(\mathbf{V} \cap \mathbf{H}^2(\Omega)\big) \times \big(L^2_{zmv}(\Omega) \cap H^1(\Omega)\big)$. *Let* $(\mathbf{u}_h, p_h) \in \mathbf{V}_{cr} \times Q_{0,h}$ *be the solution to Problem* (7.5.32). *We have the following error estimates for the discrete velocity:*

$$\|\mathbf{u} - \mathbf{u}_h\|_h \lesssim \sigma^{4d}\, h\, \|(\mathbf{u}, p)\|_\nu, \tag{7.5.45}$$

$$\text{if } \Omega \text{ is convex} \quad \|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma^{4d}\, h^2\, \|(\mathbf{u}, p)\|_\nu. \tag{7.5.46}$$

Remark 7.5.2. *Suppose that the solution to Problem* (7.2.1) *is such that* $(\mathbf{u}, p) \in \mathbf{H}^2(\Omega) \times H^1(\Omega)$. *When* Ω *is nonconvex, one can prove that* $\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma^{4d}\, h^{1+s}\, \|(\mathbf{u}, p)\|_\nu$ *for all* $s \in]\frac{1}{2}, s_\Omega[$, *where* $s_\Omega \in]\frac{1}{2}, 1[$ *depends on the reentrant faces, cf. [124, 1.A].*

Suppose that the solution to Problem (7.2.1) *is such that* $(\mathbf{u}, p) \in \mathbf{H}^{1+s}(\Omega) \times H^s(\Omega)$ *for all* $s \in]\frac{1}{2}, s_{\mathbf{f}}[$, *with* $s_{\mathbf{f}} \in]\frac{1}{2}, 1[$. *One can prove that for all* $s \in]\frac{1}{2}, s_{\mathbf{f}}[$, $\|\mathbf{u} - \mathbf{u}_h\|_h \lesssim \sigma^{2d} h^s\, \|(\mathbf{u}, p)\|_\nu$. *If* Ω *is convex, one can prove that* $\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma^{4d}\, h^{1+s}\, \|(\mathbf{u}, p)\|_\nu$. *If* Ω *is nonconvex, one can prove that* $\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma^{4d}\, h^{s+s'}\, \|(\mathbf{u}, p)\|_\nu$, $s \in]\frac{1}{2}, s_{\mathbf{f}}[$ *and* $s' \in]\frac{1}{2}, s_\Omega[$.

The proof of Theorem 7.5.3 is similar to that of [121, Theorems 3 and 4], using the properties of the Fortin operator Π_h. We detail this proof in Appendix B.2. We can then prove the convergence of the discrete pressure.

Theorem 7.5.4 (`Convergence of the discrete pressure`). *Suppose that Hypothesis 7.5.1 holds and that the solution to Problem* (7.2.1) *is such that* $(\mathbf{u}, p) \in \big(\mathbf{V} \cap \mathbf{H}^2(\Omega)\big) \times \big(L^2_{zmv}(\Omega) \cap H^1(\Omega)\big)$. *Let* $(\mathbf{u}_h, p_h) \in \mathbf{V}_{cr} \times Q_{0,h}$ *be the solution to Problem* (7.5.32). *We have the following error estimates for the discrete pressure:*

$$\|(p - p_h) - \pi_0(p - p_h)\|_{L^2(\Omega)} \lesssim \nu\, \sigma^{(4\,d-2)}\, h\, \|(\mathbf{u}, p)\|_\nu, \tag{7.5.47}$$

$$\|\pi_0(p - p_h)\|_{L^2(\Omega)} \lesssim \nu\, C_{\mathrm{div}}\, \sigma^{(4\,d+1)}\, h\, \|(\mathbf{u}, p)\|_\nu, \tag{7.5.48}$$

Remark 7.5.3. *Using Proposition 7.5.1, Equations* (7.5.47) *and* (7.5.48) *imply that*

$$\|p - p_h\|_{L^2(\Omega)} \lesssim \nu\, C_{\mathrm{div}}\, \sigma^{(4\,d+1)}\, h\, \|(\mathbf{u}, p)\|_\nu. \tag{7.5.49}$$

Let us prove Theorem 7.5.4.

Proof. Let $(\mathbf{u}_h, p_h) \in \mathcal{X}_h$ be the solution to Problem (7.5.32) with $p_h = p_{0,h} + p_{1,h}$, $(p_{0,h}, p_{1,h}) \in Q_{0,h} \times Q_{1,h}$. Integrating by parts the right-hand side in variational formulation (7.5.32), we obtain that for all $\mathbf{v}_h \in \mathbf{X}_{0,h}$:

$$(\mathbf{f}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = -\nu(\Delta\mathbf{u}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} + (\nabla\, p, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = \nu(\mathbf{u}, \mathbf{v}_h)_h - \nu \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\boldsymbol{\nabla}\, \mathbf{u} : \mathbf{n}_{f,\ell}, \mathbf{v}_h)_{\mathbf{L}^2(F_f)} + (\nabla\, p, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}. \tag{7.5.50}$$

Developping (7.5.32) with $q_h = 0$, noticing that the discrete velocity $\mathbf{u}_h \in \mathbf{V}_{cr}$, and using (7.5.50), we deduce that:

$$\nu(\mathbf{u}_h, \mathbf{v}_h)_h - (\mathrm{div}_h\, \mathbf{v}_h, p_{0,h})_{L^2(\Omega)} + (\nabla\, p_{1,h}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = \nu(\mathbf{u}, \mathbf{v}_h)_h - \nu \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\boldsymbol{\nabla}\, \mathbf{u} : \mathbf{n}_{f,\ell}, \mathbf{v}_h)_{\mathbf{L}^2(F_f)} + (\nabla\, p, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}. \tag{7.5.51}$$

Reshuffling the terms in (7.5.51), we get:

$$-(\mathrm{div}_h\,\mathbf{v}_h, p_{0,h})_{L^2(\Omega)} - (\nabla(p - p_{1,h}), \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = \varepsilon(\mathbf{u}, \mathbf{u}_h, \mathbf{v}_h), \tag{7.5.52}$$

where:

$$\varepsilon(\mathbf{u}, \mathbf{u}_h, \mathbf{v}_h) := \nu(\mathbf{u} - \mathbf{u}_h, \mathbf{v}_h)_h - \nu \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\boldsymbol{\nabla}\,\mathbf{u} : \mathbf{n}_f, \mathbf{v}_h)_{\mathbf{L}^2(F_f)}. \tag{7.5.53}$$

We can bound $\varepsilon(\mathbf{u}, \mathbf{u}_h, \mathbf{v}_h)$ using (7.5.45) and applying (7.4.2) to all the components of $\boldsymbol{\nabla}\,\mathbf{u} \in \mathbb{H}^1(\Omega)$ and $\mathbf{v}_h$, obtaining:

$$\varepsilon(\mathbf{u}, \mathbf{u}_h, \mathbf{v}_h) \lesssim \sigma^{2\,d}\, h\, \nu\, \|(\mathbf{u}, p)\|_\nu\, \|\mathbf{v}_h\|_h. \tag{7.5.54}$$

We first prove (7.5.47). Let $\mathbf{v}_h \in \mathbf{V}_{0,cr}$ cf. (7.4.8). Equation (7.5.52) now reads:

$$-(\nabla(p - p_{1,h}), \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = \varepsilon(\mathbf{u}, \mathbf{u}_h, \mathbf{v}_h). \tag{7.5.55}$$

Recall that π_1 is the L^2-orthogonal projection on $Q_{1,h}$, cf. Definition 7.5.1. Let us write (7.5.55) this way:

$$-(\nabla(\pi_1(p) - p_{1,h}), \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = \varepsilon(\mathbf{u}, \mathbf{u}_h, \mathbf{v}_h) + (\nabla(p - \pi_1(p)), \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}. \tag{7.5.56}$$

Notice that: $(\nabla(p - \pi_1(p)), \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (p - \pi_1(p), \mathbf{v}_h \cdot \mathbf{n}_{f,\ell})_{L^2(F_f)}$.
Applying (7.4.2) to all the components of $\mathbf{v}_h$ with $q = p - \pi_1(p)$, and using (7.5.43), we get:

$$\left|(\nabla(p - \pi_1(p)), \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}\right| \lesssim (\sigma + 1)\, h |p|_{H^1(\Omega)}\, \|\mathbf{v}_h\|_h. \tag{7.5.57}$$

Using (7.5.54) and (7.5.57) to bound (7.5.56), we deduce that:

$$\left|(\nabla(\pi_1(p) - p_{1,h}), \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}\right| \lesssim \sigma^{2\,d}\, h\, \nu\, \|(\mathbf{u}, p)\|_\nu\, \|\mathbf{v}_h\|_h. \tag{7.5.58}$$

Consider $\mathbf{v}_{q_{1,h}} \in \mathbf{X}^T_{0,h}$ built like $\mathbf{v}_{q_1}$ in Lemma 7.5.1 with here $q_1 = q_{1,h} := \pi_1(p) - p_{1,h}$. Using (7.5.6) and (7.5.7), it holds:

$$\|q_{1,h} - \pi_0(q_{1,h})\|_{L^2(\Omega)}\, \|\mathbf{v}_{q_{1,h}}\|_h \lesssim \sigma^{2\,(d-1)}\, (\nabla\, q_{1,h}\,,\, \mathbf{v}_{q_{1,h}})_{\mathbf{L}^2(\Omega)}.$$

Using now (7.5.58) with $\mathbf{v}_h = \mathbf{v}_{q_{1,h}}$, we deduce from the above inequality that

$$\|q_{1,h} - \pi_0(q_{1,h})\|_{L^2(\Omega)} \leq \sigma^{(4\,d-2)}\, h\, \nu\, \|(\mathbf{u}, p)\|_\nu. \tag{7.5.59}$$

Using the triangular inequality, we have:

$$\|(p - p_h) - \pi_0(p - p_h)\|_{L^2(\Omega)} \leq \|q_{1,h} - \pi_0(q_{1,h})\|_{L^2(\Omega)} + \|(p - \pi_1(p)) - \pi_0(p - \pi_1(p))\|_{L^2(\Omega)}.$$

Hence, using (7.5.59) and (7.5.44) with $q = p$, we obtain (7.5.47).

Let us prove (7.5.48). Integrating by parts the second term in (7.5.52) and moving it to the right-hand side, we get:

$$(\mathrm{div}_h\,\mathbf{v}_h, p - p_h)_{L^2(\Omega)} = \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (p - p_{1,h}, \mathbf{v}_h \cdot \mathbf{n}_{f,\ell})_{L^2(F_f)} + \varepsilon(\mathbf{u}, \mathbf{u}_h, \mathbf{v}_h). \tag{7.5.60}$$

We write: $p - p_{1,h} = (p - \pi_1(p)) + (\pi_1(p) - p_{1,h})$ where $\pi_1(p) - p_{1,h} \in \mathcal{P}^1(\mathcal{T}_h)$.
Hence, using (7.4.2) and (7.4.3), it holds, for all $\mathbf{v}_h \in \mathbf{X}_{0,h}$:

$$\left|\sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (p - p_{1,h}, \mathbf{v}_h \cdot \mathbf{n}_{f,\ell})_{L^2(F_f)}\right| \leq \sigma\, \left(h\, \|\,\nabla(p - \pi_1(p))\|_{\mathbf{L}^2(\Omega)} + \|(Id - \pi_0)(\pi_1(p) - p_{1,h})\|_{L^2(\Omega)}\right)\, \|\mathbf{v}_h\|_h. \tag{7.5.61}$$

According to (7.5.42), we have: $\|\,\nabla(p - \pi_1(p))\|_{\mathbf{L}^2(\Omega)} \leq \sigma\, \nu\, \|(\mathbf{u}, p)\|_\nu$. To bound the second term of (7.5.61), notice first that:

$$(Id - \pi_0)(\pi_1(p) - p_{1,h}) = (Id - \pi_0)(\pi_1(p) - p_h) = (Id - \pi_0)(\pi_1(p) - p) + (Id - \pi_0)(p - p_h).$$

Using a triangular inequality, we deduce that:

$$\|(Id - \pi_0)(\pi_1(p) - p_{1,h})\|_{L^2(\Omega)} \leq \|(Id - \pi_0)(\pi_1(p) - p)\|_{L^2(\Omega)} + \|(Id - \pi_0)(p - p_h)\|_{L^2(\Omega)}$$

We now use (7.5.44) and (7.5.47) to bound the first and the second term in the above equation:

$$\|(Id - \pi_0)(\pi_1(p) - p_{1,h})\|_{L^2(\Omega)} \lesssim \nu\, \sigma\, h\, \|(\mathbf{u}, p)\|_\nu + \nu\, C_{\mathrm{div}}\, \sigma^{4d}\, \|(\mathbf{u}, p)\|_\nu.$$

Finally, we have:

$$\left|\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}(p-p_{1,h},\mathbf{v}_h\cdot\mathbf{n}_{f,\ell})_{L^2(F_f)}\right|\lesssim\nu\,\sigma^{4d+1}\,h\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{v}_h\|_h. \tag{7.5.62}$$

Using (7.5.62) and (7.5.54) in (7.5.60), it holds, for all $\mathbf{v}_h\in\mathbf{X}_{0,h}$:

$$\left|(\mathrm{div}_h\,\mathbf{v}_h,p-p_h)_{L^2(\Omega)}\right|\lesssim\nu\,\sigma^{4d+1}\,h\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{v}_h\|_h. \tag{7.5.63}$$

Let $q_{0,h}=-\pi_0(p-p_h)$. Consider Equation (7.5.63) with $\mathbf{v}_{q_{0,h}}\in\mathbf{X}_{0,h}$ built like $\mathbf{v}_{q_{0,h}}$ in Lemma 7.4.1. We have:

$$\nu^{-1}\|\pi_0(p-p_h)\|^2_{L^2(\Omega)}=(\mathrm{div}_h\,\mathbf{v}_{q_{0,h}},p-p_h)_{L^2(\Omega)}\text{ and }\|\mathbf{v}_{q_{0,h}}\|_h\le\nu^{-1}\,C_{\mathrm{div}}\|\pi_0(p-p_h)\|_{L^2(\Omega)}.$$

Using this in (7.5.63), we deduce that:

$$\nu^{-1}\|\pi_0(p-p_h)\|^2_{L^2(\Omega)}\lesssim\nu\,\sigma^{(4d+1)}\,h\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{v}_{\pi_0(q)}\|_h\lesssim\sigma^{(4d+1)}\,C_{\mathrm{div}}\,h\,\|(\mathbf{u},p)\|_\nu\,\|\pi_0(p-p_h)\|_{L^2(\Omega)}.$$

Simplifying by $\nu^{-1}\,\|\pi_0(p-p_h)\|_{L^2(\Omega)}$, we obtain (7.5.48). ⋄

Remark 7.5.4. *Consider inhomogeneous boundary conditions:* $\mathbf{u}_{|\partial\Omega}=\mathbf{g}\in\mathbf{H}^{\frac{1}{2}}(\partial\Omega)$ *such that* $\int_{\partial\Omega}\mathbf{g}\cdot\mathbf{n}_{|\partial\Omega}=0$. *One can proceed as explained in [121, Section 7]. Let us set* $\gamma_{\partial\Omega}(\mathbf{X}_h):=\{\mathbf{v}_{h|\partial\Omega}$ *such that* $\mathbf{v}_h\in\mathbf{X}_h\}$ *and:*

$$\mathbf{G}_h:=\left\{\mathbf{v}_h\in\gamma_{\partial\Omega}(\mathbf{X}_h)\ \textit{and}\ \int_{\partial\Omega}\mathbf{v}_h\cdot\mathbf{n}_{|\partial\Omega}=0\right\},\quad\tilde{\mathbf{G}}_h:=\{\mathbf{v}_h\in\gamma_{\partial\Omega}(\mathbf{X}_h)\ \textit{and}\ \forall q_h\in Q_h,\,b_h(\mathbf{v}_h,q_h)=0\}.$$

According to [121, Lemma 9], $\mathbf{G}_h=\{\mathbf{v}_h\in\gamma_{\partial\Omega}(\mathbf{X}_h)$ *and* $\forall q_h\in Q_{0,h},\,(\mathrm{div}_h\,\mathbf{v}_h,q_h)=0\}$. *Hence,* $\tilde{\mathbf{G}}_h\subset\mathbf{G}_h$. *Notice that we can extend the definition of our Fortin operator* Π_h *from* $\mathbf{H}^1(\Omega)$ *to* $\mathbf{X}_h$ *(cf. Proposition 7.5.3). Alternatively, one can prove discrete T-coercivity of the bilinear form* $a_h(\cdot,\cdot)$, *and use the second Strang's Lemma [148, §27.4.2] to compute error estimates. We won't go into more detail here, as it's fairly technical.*

We proved the convergence of the $\mathbf{P}^1_{nc}\times(P^0+P^1)$ finite element. In the next section, we prove that we can obtain optimal error estimates in $2D$.

7.5.5 Optimal error estimates in $2D$

We consider here $\Omega\subset\mathbb{R}^2$. In order to exhibit some convergence properties of the $\mathbf{P}^1_{nc}\times(P^0+P^1)$ finite element method, we need the two next Lemmas:

Lemma 7.5.2. *Let* $\mathbf{v}_h\in\mathbf{V}_{cr}$. *Then for all* $q_{2,h}\in Q_{2,h}$, $(\nabla\,q_{2,h},\mathbf{v}_h)_{\mathbf{L}^2(\Omega)}=0$.

Proof. Integrating by parts and using Proposition 7.5.2, we get that for all $q_{2,h}\in Q_{2,h}$ and for all $q_{1,h}\in Q_{1,h}$:

$$\sum_{\ell\in\mathcal{I}_K}(\nabla\,q_{2,h},\mathbf{v}_h)_{\mathbf{L}^2(K_\ell)}=-\sum_{f\in\mathcal{I}_F}\int_{F_f}q_{2,h}[\mathbf{v}_h]_f\cdot\mathbf{n}_f=-\sum_{f\in\mathcal{I}_F}\int_{F_f}(q_{2,h}-q_{1,h})[\mathbf{v}_h]_f\cdot\mathbf{n}_f.$$

Hence, we obtain for all $q_{2,h}\in Q_{2,h}$ and for all $q_{1,h}\in Q_{1,h}$:

$$\begin{aligned}\sum_{\ell\in\mathcal{I}_K}(\nabla\,q_{2,h},\mathbf{v}_h)_{\mathbf{L}^2(K_\ell)}&=-\sum_{f\in\mathcal{I}_F}\int_{F_f}(q_{2,h}-q_{1,h})[\mathbf{v}_h-\underline{\mathbf{v}}_{h,f}]_f\cdot\mathbf{n}_f\\&=-\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}\int_{F_f}(q_{2,h}-q_{1,h})(\mathbf{v}_h-\underline{\mathbf{v}}_{h,f})\cdot\mathbf{n}_{f,\ell}.\end{aligned}$$

Let us build $q_{1,h}\in Q_{1,h}$ such that for all $i\in\mathcal{I}_S$, $q_{1,h}(S_i)=q_{2,h}(S_i)$. Then, for all $f\in\mathcal{I}_F$, the polynomial $(q_{2,h}-q_{1,h})(\mathbf{v}_h-\underline{\mathbf{v}}_{h,f})$, which is of order 3, vanishes at the vertices and in the middle of F_f. Using the Simpson's rule which is exact for order 3 polynomials, we obtain that

$$\int_{F_f}(q_{2,h}-q_{1,h})(\mathbf{v}_h-\underline{\mathbf{v}}_{h,f})\cdot\mathbf{n}_{f,\ell}=0.$$

⋄

Lemma 7.5.3. *Let $\phi \in H^3(\Omega) \cap L^2_{zmv}(\Omega)$ such that $|\phi|_{H^3(\Omega)} \neq 0$. Then, for all $\mathbf{v}_h \in \mathbf{V}_{cr}$:*

$$(\nabla \phi, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} \lesssim \sigma\, h^2\, |\phi|_{H^3(\Omega)}\, \|\mathbf{v}_h\|_{\mathbf{L}^2(\Omega)}. \tag{7.5.64}$$

Proof. Let $\mathcal{I}_{2,h}$ be the P^2-interpolation operator on $\mathcal{T}_h$. Due to Lemma 7.5.2, for all $\mathbf{v} \in \mathbf{V}_{cr}$, we have: $(\nabla \phi, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = (\nabla(\phi - \mathcal{I}_{2,h}(\phi)), \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}$. Using Cauchy-Schwarz inequality and [147, Theorem 11.13], we get that:

$$\left|(\nabla \phi, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)}\right| \leq \|\nabla(\phi - \mathcal{I}_{2,h}(\phi))\|_{\mathbf{L}^2(\Omega)}\, \|\mathbf{v}_h\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma\, h^2\, |\phi|_{H^3(\Omega)}\, \|\mathbf{v}_h\|_{\mathbf{L}^2(\Omega)}.$$

◇

We can now prove the following result.

Theorem 7.5.5 (`Pressure robustness`)**.** *Suppose that Ω is convex. Let $\phi \in H^3(\Omega) \cap L^2_{zmv}(\Omega)$ such that $|\phi|_{H^3(\Omega)} \neq 0$. Consider Problem (7.2.1) with $\mathbf{f} = \nabla \phi$. Then the solution to Problem (7.2.1) is such that: $(\mathbf{u}, p) = (0, \phi)$. Let $(\mathbf{u}_h, p_h)$ be the solution to Problem (7.5.32). It is such that:*

$$\nu\, \|\mathbf{u}_h\|_h \lesssim \sigma^2\, h^3\, |\phi|_{H^3(\Omega)}, \tag{7.5.65}$$

$$\nu\, \|\mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma^{(2\,d+1)}\, h^4\, |\phi|_{H^3(\Omega)} \tag{7.5.66}$$

$$\|(\phi - p_h) - \pi_0(\phi - p_h)\|_{L^2(\Omega)} \lesssim \sigma^{2\,d}\, h^2\, |\phi|_{H^2(\Omega)}, \tag{7.5.67}$$

$$\|\pi_0(\phi - p_h)\|_{L^2(\Omega)} \lesssim C_{\rm div}\, \sigma^{2\,d+1}\, h^2\, |\phi|_{H^2(\Omega)}. \tag{7.5.68}$$

Proof. Setting $\mathbf{v}_h = \mathbf{u}_h \in \mathbf{V}_{cr}$ as test-function in Problem (7.5.32), using Lemma 7.5.2, integrating by parts, we have:

$$\nu\, \|\mathbf{u}_h\|_h^2 = (\nabla(\phi - \mathcal{I}_{2,h}(\phi)), \mathbf{u}_h)_{\mathbf{L}^2(\Omega)} = \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\phi - \mathcal{I}_{2,h}(\phi), \mathbf{u}_h \cdot \mathbf{n}_{f,\ell})_{L^2(F_f)}.$$

Hence, using (7.4.2) and [147, Theorem 11.13], we obtain:

$$\nu\, \|\mathbf{u}_h\|_h^2 \lesssim \sigma\, h\, \|\nabla(\phi - \mathcal{I}_{2,h}(\phi))\|_{\mathbf{L}^2(\Omega)}\, \|\mathbf{u}_h\|_h \lesssim \sigma^2\, h^3\, |\phi|_{H^3(\Omega)}\, \|\mathbf{u}_h\|_h.$$

We deduce (7.5.65). Using this estimate in the proof of (7.5.46), we obtain (7.5.66). The proofs of estimates (7.5.67)-(7.5.68) are similar to the proofs of estimates (7.5.47)-(7.5.48), using (7.5.65)-(7.5.66) and the fact that since $\phi \in H^3(\Omega)$, $\|\nabla(\phi - \pi_1(\phi))\|_{\mathbf{L}^2(\Omega)} \lesssim \sigma\, h\, |\phi|_{H^2(\Omega)}$. ◇

Remark 7.5.5. *In the case of nonhomogeneous Dirichlet boundary conditions, we consider Problem B.4.2 given in Appendix B.4. We can use T-coercivity to state the well-posedness of the bilinear form $a_h(\cdot,\cdot)$ defined (7.5.31). The proof of convergence relies on the second Strang's Lemma [148, Lemma 27.15].*

Numerical results illustrating the optimal convergence rates are given in §7.5.6.

7.5.6 Numerical illustration [A13, W1]

To get $p_h \in L^2_{zmv}(\Omega)$, we eliminate a degree of freedom in $\mathcal{P}^0_{disc}(\mathcal{T}_h)$. We compute the discrete errors values $\varepsilon^\nu_0(\mathbf{u}_h)$, $\varepsilon^\nu_1(\mathbf{u}_h)$ and $\varepsilon^\nu_0(p_h)$ (cf. definition 7.4.1). Let $\tau_{0,\mathbf{u}}$, $\tau_{1,\mathbf{u}}$ and τ_p be the averaged convergence rates of $\varepsilon^\nu_0(\mathbf{u}_h)$, $\varepsilon^\nu_1(\mathbf{u}_h)$ and $\varepsilon^\nu_0(p_h)$.

We first consider a $2D$ case illustrating Theorem 7.5.5, with the prescribed solution $\mathbf{u} = 0$, $\mathbf{f} = \nabla \psi$ with $\psi(x, y) = \exp(-10(1-x+2y))$. Figure 7.2 shows $\varepsilon^\nu_0(\mathbf{u}_h)$, $\varepsilon^\nu_1(\mathbf{u}_h)$ and $\varepsilon^\nu_0(p_h)$ against h, for $\nu = 10^{-2}$, with P^0 (left) or $P^0 + P^1$ (right) discrete pressures, comparing the $\mathbf{P}^1_{nc} \times P^0$ [121, Example 4] and $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ schemes. Table 7.5 reports observed convergence rates, which are better for the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ scheme.

Second, we consider stationary Navier-Stokes equations set in $\Omega = (0,1)^3$: Find $(\mathbf{u}, p)$ such that $-\nu\, \boldsymbol{\Delta}\, \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p = \mathbf{f}$ and $\operatorname{div} \mathbf{u} = 0$ in Ω, with nonhomogeneous Dirichlet boundary conditions. Computations are done with the TrioCFD code [11, A13] with the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ scheme. The convection term is discretized using the MUSCL scheme [11, A13]. We study the prescribed solution

$\mathbf{u} = (y-z, z-x, x-y)^T$, $p = \frac{1}{2}(x^2+y^2+z^2) - xy - xz - yz - 1/4$ so that $\mathbf{f} = 0$. Figure 7.3 shows $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right) against h, for $\nu = 10^{-1}; 10^{-2}; 10^{-3}$. Table 7.6 reports the convergence rates $\tau_{0,\mathbf{u}}$ (left), and τ_p (right).

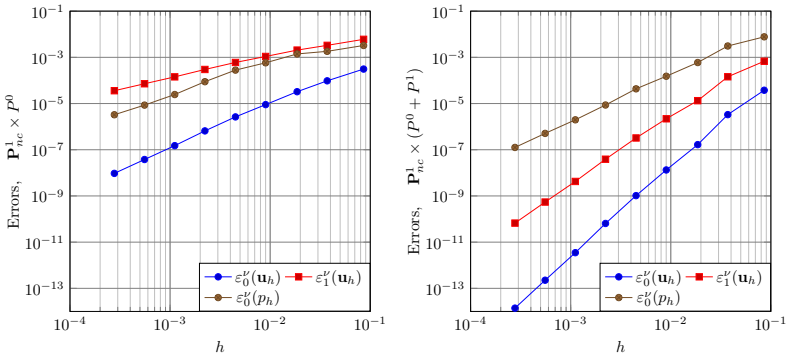


Figure 7.2: $2D$ case. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$, $\varepsilon_1^\nu(\mathbf{u}_h)$ and $\varepsilon_0^\nu(p_h)$ for $\nu = 10^{-2}$.

	P^0	$P^0 + P^1$
$\tau_{0,\mathbf{u}}$	2.02	4.03

	P^0	$P^0 + P^1$
$\tau_{1,\mathbf{u}}$	1.01	3.07

	P^0	$P^0 + P^1$
τ_p	1.60	2.10

Table 7.5: $2D$ case, averaged convergence rates on the last five meshes.

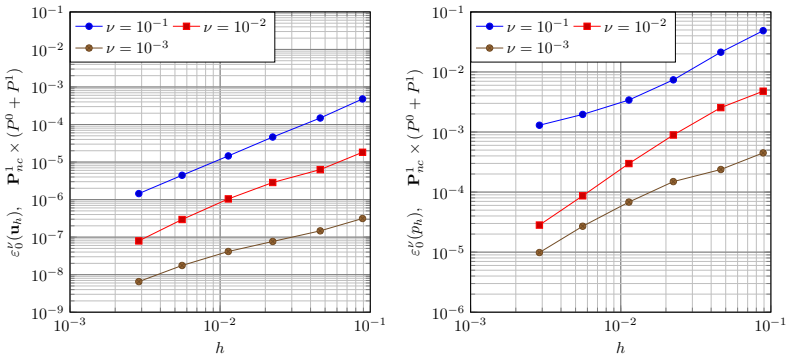


Figure 7.3: $3D$ case. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ and $\varepsilon_0^\nu(p_h)$ for $\nu = 10^{-1}$, $\nu = 10^{-2}$ or $\nu = 10^{-3}$.

More numerical results are available in [175, 153, 11, 165]. The $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ scheme shows good approximation properties and should therefore be suitable for solving the Navier-Stokes equations in the context of industrial simulation. More generally, it seems interesting to design and study discrete pressure enrichment for higher order nonconforming Crouzeix-Raviart mixed FEM [227, 228].

Another strategy to improve the accuracy of the $\mathbf{P}^1_{nc} \times P^0$ finite elements may be to modify the discretization of the gradient of the pressure, as we will see next.

ν	10^{-1}	10^{-2}	10^{-3}
$\tau_{0,\mathbf{u}}$	1.68	1.88	1.41

ν	10^{-1}	10^{-2}	10^{-3}
τ_p	0.70	1.72	1.35

Table 7.6: $3D$ case, averaged convergence rates on the last three meshes.

7.6 The $\mathbf{P}^1_{nc} \times P^0_{Mps}$ scheme [P3, P4, 210]

We propose in this Section to use the symmetric MPFA scheme (where MPFA stands for multi-points flux approximation) to discretize the pressure gradient term in (7.2.1), in the case of a simplicial triangulation. This scheme is part of the gradient scheme formalism and the resulting diffusion operator is consistent with the hypothesis given by [136, Theorem 12.5]. The discrete pressure space remains $Q_{0,h}$. We call this new scheme the $\mathbf{P}^1_{nc} \times P^0_{Mps}$ scheme. Our aim is to design a scheme with only P^0 unknowns for the pressure, and which would be more precise than the $\mathbf{P}^1_{nc} \times P^0$ scheme.

7.6.1 Description of the scheme

The $\mathbf{P}^1_{nc} \times P^0_{Mps}$ scheme has been implemented in $2D$ on an Octave model and in $3D$ in the TrioCFD code. We refer to [210, §5.2] for the $3D$ case. Another description of this scheme, based on the relation between nonconforming finite elements and the symmetric MPFA scheme, is proposed in Appendix B.3.

Consider the $2D$ case. To design the scheme, initially proposed by Le Potier in [186, 187], we start by splitting the triangles into three quadrangles, connecting the barycentre of the triangle to the midpoint of each edges. Considering some $q_h \in Q_{0,h}$, we will calculate an affine approximation of $Q_{0,h}$ on each quadrangle. To do so, we need to add temporary auxiliary unknowns located on the third of the edges. Let us introduce some notations.

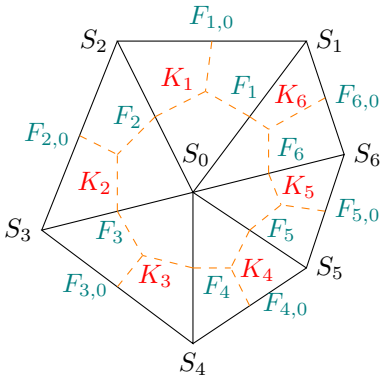
Let $j \in \mathcal{I}_S$. We define the macro-element $\mathcal{M}_j$ such that $\overline{\mathcal{M}}_j := \bigcap_{\ell \in \mathcal{I}_{K,j}} \overline{K}_\ell$. Let renumber the vertices so that: $S_0 = S_j$, $\mathcal{I}_{S,0} = \{1, \dots, N_{S,0}\}$ and for all $i \in \mathcal{I}_{S,0}$, $S_i S_{i+1} \in \mathcal{F}_h$ (setting $S_{N_{S,0}+1} = S_{N_{S,0}}$). For $i \in \mathcal{I}_{S,0}$ we have:

- The triangle $K_i = S_0 S_i S_{i+1}$, and we call its barycentre G_i.
- The edge $F_i = S_0 S_i$, of barycentre M_i.
- $F_{i,0}$ the edge opposite to S_0 in K_i.
- The half-edges $\tilde{F}_i = S_0 M_i$.
- The quadrangle $Q_i = S_0 M_i G_i M_{i+1}$.

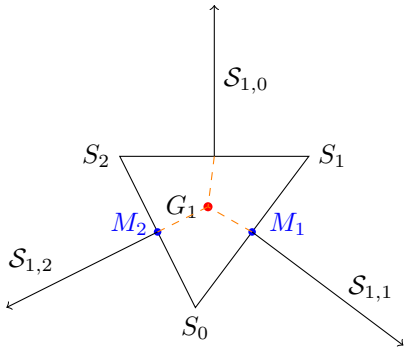
For i, $j \in \mathcal{I}_{S,0}$, we denote by $\mathcal{S}_{i,j}$ the normal vector outgoing of K_i in F_j and of norm $|F_j|$. For $i \in \mathcal{I}_{S,0}$, we call $\mathcal{S}_{i,0}$ the normal vector outgoing of K_i at $F_{i,0}$. In Figure 7.4(a), we represent $\mathcal{M}_0$ in case $N_{S,0} = 6$ and in Figure 7.4(b), we represent the triangle K_1 with the vectors $(\mathcal{S}_{1,j})_{j=0}^2$ and its barycentre G_1. Let $q_h \in Q_{0,h}$. We set $q_{h|K_\ell} := \overline{q}_\ell$.

Suppose first that $S_0 \in \Omega$, as in Figure 7.5. We build a piecewise affine approximation of q_h on each quadrangle $(Q_i)_{i \in \mathcal{I}_{S,0}}$, called $\tilde{q}_h$. To do so, we introduce discrete pressure values $(\tilde{q}_i)_{i \in \mathcal{I}_{S,0}}$ on the thirds of the inner edges of $\mathcal{M}_0$, cf. Figure 7.6(b). For all $j \in \mathcal{I}_{S,i}$, we define $\mathcal{G}_{0,i}(q_h) := \nabla \tilde{q}_{h|Q_i}$, using an integration by parts as it is done in [186, §3] and [188, §1.1.1]:

$$|Q_i|\, \mathcal{G}_{0,i} = \int_{Q_i} \mathcal{G}_{0,i}(q_h) = \int_{\partial Q_i} \tilde{q}_h \mathbf{n}_{|\partial Q_i} = \tilde{q}_i \frac{\mathcal{S}_{i,i}}{d} + \tilde{q}_{i+1} \frac{\mathcal{S}_{i+1,i}}{d} + \overline{q}_i \left(-\frac{\mathcal{S}_{i,i}}{d} - \frac{\mathcal{S}_{i+1,i}}{d} \right).$$

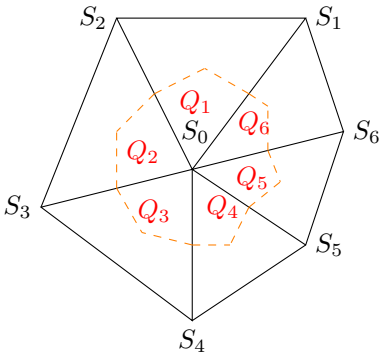


(a) Macro-element $\mathcal{M}_0 = S_1 \dots S_6$

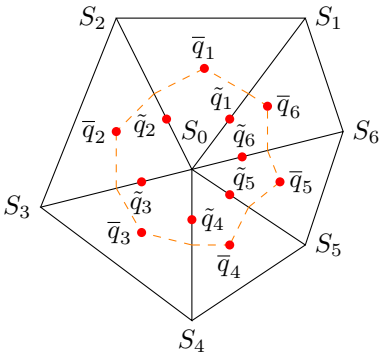


(b) Triangle $K_1 = S_0S_1S_2$

Figure 7.4: Notations, $N_{S,0} = 6$ and $j \in \mathcal{I}_S^i$.

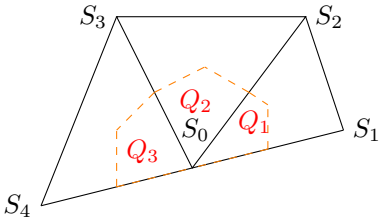


(a) Quadrangles $(Q_i)_{i=1}^6$

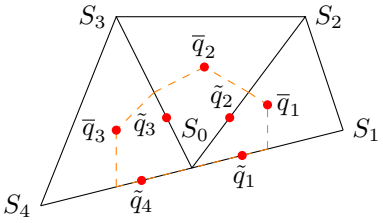


(b) Discrete pressures $(\overline{q}_i, \tilde{q}_i)_{i=1}^6$

Figure 7.5: MPFA scheme, $N_{S,0} = 6$ and $j \in \mathcal{I}_S^i$.



(a) Quadrangles $(Q_i)_{i=1}^4$



(b) Discrete pressures $(\overline{q}_i, \tilde{q}_i)_{i=1}^4$

Figure 7.6: MPFA scheme, $N_{S,0} = 4$ and $j \in \mathcal{I}_S^b$.

Noticing that $Q_i = |K_i|/d + 1$, we have:

$$\mathcal{G}_{0,i}(q_h) = \frac{1}{|Q_i|}\left((\tilde{q}_i - \overline{q}_i)\frac{\mathcal{S}_{i,i}}{d} + (\tilde{q}_{i+1} - \overline{q}_i)\frac{\mathcal{S}_{i+1,i}}{d}\right) = \frac{d+1}{d|K_i|}\left(\tilde{q}_i\mathcal{S}_{i,i} + \tilde{q}_{i+1}\mathcal{S}_{i+1,i} + \overline{q}_i\mathcal{S}_{0,i}\right) \tag{7.6.1}$$

In order to (weakly) preserve the continuity of the normal component of the flux across the inner edges of $\mathcal{M}_0$, we write:

$$\forall i \in \mathcal{I}_S, \quad \mathcal{G}_{0,i} \cdot \mathcal{S}_{i,i+1} + \mathcal{G}_{0,i+1} \cdot \mathcal{S}_{i+1,i+1} = 0. \tag{7.6.2}$$

Having N_{S_0} equations and as many unknowns, we can evaluate the auxiliary discrete pressure values $(\tilde{q}_i)_{i\in\mathcal{I}_{S,0}}$ with data $(\overline{q}_i)_{i\in\mathcal{I}_{S,0}}$, and then express pressure gradients $(\mathcal{G}_{0,i}(q_h))_{i\in\mathcal{I}_{S,0}}$ (7.6.1).

Suppose now that $S_0 \in \partial\Omega$, as in Figure 7.6. We need to explicit the auxiliary discrete pressure values located on $\partial\Omega$ (i.e. $\tilde{q}_1$ and $\tilde{q}_4$ in Figure 7.6(b)). Recall that $\nabla p = \mathbf{f} + \nu\Delta\mathbf{u}$. Suppose that $\mathbf{f}\cdot\mathbf{n}_{|\partial\Omega} \in L^1(\partial\Omega)$. Noticing that ν may be very small, we impose that for all $i \in \mathcal{I}_{S,0}$ such that $F_i \subset \partial\Omega$:

$$\int_{\tilde{F}_i} \mathcal{G}_i \cdot \mathbf{n}_{|\tilde{F}_i} = \int_{\tilde{F}_i} \mathbf{f}\cdot\mathbf{n}_{|\tilde{F}_i}. \tag{7.6.3}$$

Again, the auxiliary discrete pressure values solve a well posed linear system. They can be written with the P^0 values $(\overline{q}_i)_{i\in\mathcal{I}_{S,0}}$ and we can explicitly express the gradients $(\mathcal{G}_{0,i}(q_h))_{i\in\mathcal{I}_{S,0}}$.

For $i \in \mathcal{I}_S$, we let $(Q_{i,j})_{j\in\mathcal{I}_{S,i}}$ be the set of quadrangles built around S_i, and we call $\mathcal{Q}_h$ the mesh of all the quadrangles $\mathcal{Q}_h := (\,(Q_{i,j})_{j\in\mathcal{I}_{S,i}}\,)_{i\in\mathcal{I}_S}$. Let $q_h \in Q_{0,h}$ and $i \in \mathcal{I}_S$. In the macro-element $\mathcal{M}_i$, we call $\mathcal{G}_{i,j}(q_h)$ the local reconstructed gradient of q_h. We now define the MPFA gradient reconstruction as the operator $\mathcal{G}_h : Q_h \to P^0(\mathcal{Q}_h)$ such that for all $q_h \in Q_h$:

$$\forall i \in \mathcal{I}_S,\ \forall j \in \mathcal{I}_{S,i},\ \mathcal{G}_h(q_h)_{|Q_{i,j}} = \mathcal{G}_{i,j}(q_{h|\mathcal{M}_i}). \tag{7.6.4}$$

Alternatively, one can enhance the space of discrete pressures, adding the auxiliary unknowns on the boundary as degrees of freedom. The symmetric MPFA scheme applied to the diffusion operator exhibits some properties:

Proposition 7.6.1 (`[186, Proposition 2 and 3], [197, Theorem 3.2]`). *The consistent approximation of fluxes in the symmetric MPFA scheme is achieved using triangles and $p \in \mathcal{C}^2(\Omega)$. Additionally, exact approximation of gradients for affine functions is obtained by selecting the auxiliary pressure unknowns at the thirds of the edges. Furthermore, the symmetric MPFA scheme exhibits properties of consistency, coercivity, and convergence.*

The discretization of (7.2.1) with $g = 0$ using the MPFA scheme to discretize the pressure gradient reads: Find $(\mathbf{u}_h, p_h) \in \mathcal{X}_h$ such that for all $(\mathbf{v}_h, q_h) \in \mathcal{X}_h$

$$\nu(\mathbf{u}_h,\mathbf{v}_h)_h + (\mathbf{v}_h, \mathcal{G}_h(q_h))_{\mathbf{L}^2(\Omega)} - (\mathrm{div}_h\,\mathbf{v}_h, q_h)_{L^2(\Omega)} = \ell_{\mathbf{f}}(\,(\mathbf{v}_h, q_h)\,), \tag{7.6.5}$$

where $\mathcal{X}_h = \mathbf{X}_{0,h} \times Q_{0,h}$ and $\ell_{\mathbf{f}}(\cdot)$ is defined in (7.4.9).

We didn't prove yet that problem (7.6.5) is well-posed. Nevertheless, we obtained interesting numerical results.

7.6.2 Numerical results [P3]

We give some $2D$ numerical results which compare the $\mathbf{P}^1_{nc} \times P^0_{Mps}$ scheme to the $\mathbf{P}^1_{nc} \times P^0$ and $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ schemes. Computation are done with an Octave digital mockup.

Consider problem (7.2.1) set in $\Omega = (0,1)^2$ with prescribed solution $(\mathbf{u}, p) = (0, \varphi)$. When φ is an affine function, we obtain $\mathbf{u}_h = 0$ up to machine precision for both the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ and $\mathbf{P}^1_{nc} \times P^0_{Mps}$ scheme. When φ is some quadratic function, we obtain $\mathbf{u}_h = 0$ up to machine precision for the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ scheme, cf. Lemma 7.5.2.

We consider problem (7.2.1) with prescribed solution $(\mathbf{u}, p) = (0, \sin(2\pi x)\,\sin(2\pi y))$. In Figure 7.7, we plot $\varepsilon^\nu_0(\mathbf{u}_h)$ and $\varepsilon^\nu_0(p_h)$ against the mesh step h in the logarithmic scale. Notice that in that case, these errors don't depend on ν. The $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ and $\mathbf{P}^1_{nc} \times P^0_{Mps}$ schemes are more accurate than the $\mathbf{P}^1_{nc} \times P^0$ scheme.

Second, we consider problem (7.2.1) with $\nu = 10^{-6}$ and prescribed solution

$$\mathbf{u} = \begin{pmatrix} (\cos(2\pi x) - 1)\,\sin(2\pi y) \\ -\cos(2\pi y) + 1)\,\sin(2\pi x) \end{pmatrix}, \quad p = \sin(2\pi x)\,\sin(2\pi y)). \tag{7.6.6}$$

In Figure 7.8, we plot $\varepsilon^\nu_0(\mathbf{u}_h)$ and $\varepsilon^\nu_0(p_h)$ against the mesh step h in the logarithmic scale. Again, the

$\mathbf{P}^1_{nc} \times (P^0 + P^1)$ and $\mathbf{P}^1_{nc} \times P^0_{Mps}$ schemes are more accurate than the $\mathbf{P}^1_{nc} \times P^0$ scheme. The convergence rates are similar to that of the first test.

Other numerical results are available in [P3, P4, 210], in particular in $3D$ for the Navier-Stokes equations. In [210], several variants of this scheme are described, in particular for the discretization of the boundary auxiliary unknowns. This method delivers promising results and it could be a good compromise between the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ and the $\mathbf{P}^1_{nc} \times P^0$ schemes. Next, we intend to prove that the (7.6.5) problem is well posed and to validate the scheme using industrial benchmarks.

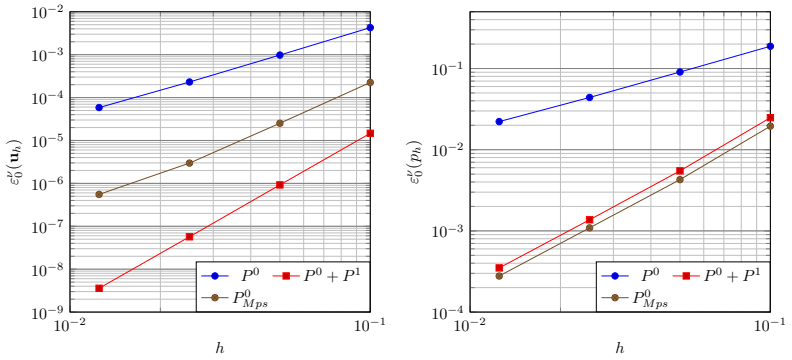


Figure 7.7: First test. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

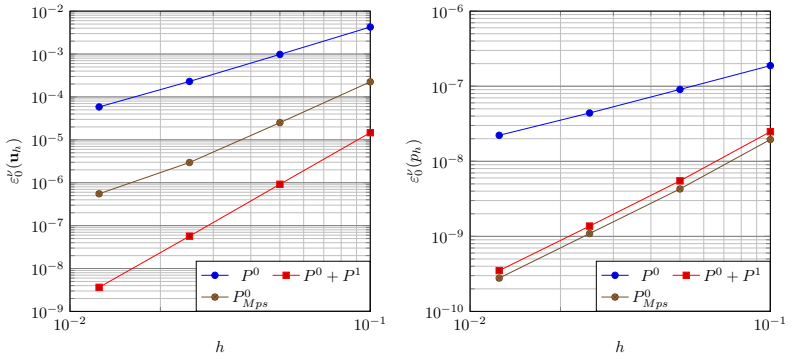


Figure 7.8: Second test. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

Test	τ	$\mathbf{P}^1_{nc} \times P^0$	$\mathbf{P}^1_{nc} \times (P^0 + P^1)$	$\mathbf{P}^1_{nc} \times P^0_{Mps}$
Case 1	$\tau_{\mathbf{u}}$	2.06	4.00	2.89
	τ_p	1.03	2.05	2.04
Case 2	$\tau_{\mathbf{u}}$	2.06	3.99	2.89
	τ_p	1.03	2.05	2.04

Table 7.7: Average convergence rates.

Comparing the error estimates with respect to the mesh size, the $\mathbf{P}^1_{nc} \times P^0_{Mps}$ scheme is more effective than $\mathbf{P}^1_{nc} \times P^0$ scheme, but it isn't as effective as the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$. Further investigation is required in the $3D$ case. Notice that the MPFA scheme can be used on triangulations with general meshes. Thus, one could expect to extend the $\mathbf{P}^1_{nc} \times P^0_{Mps}$ scheme with the lowest order of the Hybrid High Order approximation for the velocity, cf. [122]. Indeed, this discretization, which is adapted polytopal meshes, corresponds to the non-conforming discretization in the case of a simplicial triangulation.

In the next Section, we propose a new approach to improve the $\mathbf{P}^1_{nc} \times P^0$ discretization, using explicit T-coercivity.

7.7 Explicit T-coercivity [A5]

We solve the classical incompressible Stokes model, with $g = 0$. Let $\lambda > \frac{1}{4}(C_{\rm div})^2$. In remark 7.2.1, we introduced the operator $T_\lambda \in \mathcal{L}(\mathcal{X})$ defined by:

$$\left\{ \begin{array}{rcl} T_\lambda : \mathcal{X} & \to \mathbb{R} & \\ (\mathbf{v}, q) & \mapsto & (\lambda\, \mathbf{v} - \nu^{-1}\tilde{\mathbf{v}}_q, -\lambda\, q) \end{array} \right. , \tag{7.7.1}$$

where $\tilde{\mathbf{v}}_q \in \mathbf{V}^\perp$ is given by (4.3.1): $\operatorname{div} \tilde{\mathbf{v}}_q = q$ and $\|\tilde{\mathbf{v}}_q\|_{\mathbf{H}^1_0(\Omega)} \leq C_{\rm div}\, \|q\|_{L^2\Omega)}$. We now write the variational formulation (7.2.4) with test function $T_\lambda\,(\,(\mathbf{v}, q)\,)$ instead of $(\mathbf{v}, q)$. Let us define $\tilde{a}_\lambda\,((\mathbf{u}', p'), (\mathbf{v}, q)) = a\,((\mathbf{u}', p'), T(\mathbf{v}, q))$. We have:

$$\begin{array}{rcl} \tilde{a}_\lambda\,((\mathbf{u}', p'), (\mathbf{v}, q)) & = & \nu\,\lambda\,(\mathbf{u}', \mathbf{v})_{\mathbf{H}^1_0(\Omega)} - (\mathbf{u}', \tilde{\mathbf{v}}_q)_{\mathbf{H}^1_0(\Omega)} \\ & & -\lambda\,(p', \operatorname{div} \mathbf{v})_{L^2(\Omega)} + \nu^{-1}\,(p', q)_{L^2(\Omega)} \\ & & +\lambda\,(q, \operatorname{div} \mathbf{u}')_{L^2(\Omega)}. \end{array} \tag{7.7.2}$$

The term with $\tilde{\mathbf{v}}_q$ requires some knowledge of the nonlocal right inverse of the divergence operator. We will see next that it can be removed. According to Prop. 7.2.2, we have the

Proposition 7.7.1. *The bilinear form* (7.7.2) *is coercive.*

Introducing $\ell_\lambda(\,(\mathbf{v}, q)\,) := \ell(\,T(\mathbf{v}, q)\,)$, we can propose a first new variational formulation to Problem $(7.2.1)_H$, which reads:

$$\text{Find } (\mathbf{u}, p) \in \mathcal{X} \text{ s.t.} \quad \tilde{a}_\lambda\,((\mathbf{u}, p), (\mathbf{v}, q)) = \ell_\lambda(\,(\mathbf{v}, q)\,) \quad \forall (\mathbf{v}, q) \in \mathcal{X}. \tag{7.7.3}$$

Using the bijectivity of T, it is obvious that (7.7.3) is equivalent to (7.2.4), so well-posedness of (7.7.3) is a direct consequence of the well-posedness of (7.2.4), and vice versa.

Let us derive an "explicit" expression of $\ell_\lambda((\mathbf{v}, q))$, which will prove useful later on. We recall that we have in the sense of distributions:

$$-\,\boldsymbol{\Delta}(\cdot) = \mathbf{curl}\,\mathbf{curl}(\cdot) - \nabla \operatorname{div}(\cdot), \tag{7.7.4}$$

so that

$$\forall \mathbf{f}' \in \mathbf{H}^{-1}(\Omega), \quad \exists!(z_{\mathbf{f}'}, \mathbf{w}_{\mathbf{f}'}) \in L^2_{zmv}(\Omega) \times \mathbf{V} \,|\, \mathbf{f}' = \nabla\, z_{\mathbf{f}'} - \boldsymbol{\Delta}\, \mathbf{w}_{\mathbf{f}'} = \nabla\, z_{\mathbf{f}'} + \mathbf{curl}\,\mathbf{curl}\,\mathbf{w}_{\mathbf{f}'}, \tag{7.7.5}$$

where $(\mathbf{w}_{\mathbf{f}'}, z_{\mathbf{f}'})$ solves $(7.2.1)_H$ with $\nu = 1$ and data $\mathbf{f} = \mathbf{f}'$ and $g = 0$.

Proposition 7.7.2. *Let $\mathbf{f}' \in \mathbf{H}^{-1}(\Omega)$. Given $q \in L^2_{zmv}(\Omega)$, let $\tilde{\mathbf{v}}_q \in \mathbf{V}^\perp$ be defined by (4.3.1). We have:*

$$\langle \mathbf{f}', \tilde{\mathbf{v}}_q \rangle_{\mathbf{H}^1_0(\Omega)} = -(z_{\mathbf{f}'}, q)_{L^2(\Omega)}. \tag{7.7.6}$$

Proof. Let $\mathbf{f}'$ be decomposed as in (7.7.5). On the one hand, integrating by parts twice and using (7.7.4), we get:

$$\begin{array}{rcl} -\langle \mathbf{curl}\,\mathbf{curl}\,\mathbf{w}_{\mathbf{f}'}, \tilde{\mathbf{v}}_q \rangle_{\mathbf{H}^1_0(\Omega)} & = & -\langle \mathbf{curl}\,\mathbf{curl}\,\tilde{\mathbf{v}}_q, \mathbf{w}_{\mathbf{f}'} \rangle_{\mathbf{H}^1_0(\Omega)} = \langle \boldsymbol{\Delta}\,\tilde{\mathbf{v}}_q, \mathbf{w}_{\mathbf{f}'} \rangle_{\mathbf{H}^1_0(\Omega)} - \langle \nabla\, q, \mathbf{w}_{\mathbf{f}'} \rangle_{\mathbf{H}^1_0(\Omega)}, \\ & = & -(\tilde{\mathbf{v}}_q, \mathbf{w}_{\mathbf{f}'})_{\mathbf{H}^1_0(\Omega)} + (q, \operatorname{div} \mathbf{w}_{\mathbf{f}'})_{L^2(\Omega)} = 0, \end{array}$$

resp. since $\operatorname{div} \tilde{\mathbf{v}}_q = q$, $\tilde{\mathbf{v}}_q \in \mathbf{V}^\perp$, $\mathbf{w}_{\mathbf{f}'} \in \mathbf{V}$, and $\operatorname{div} \mathbf{w}_{\mathbf{f}'} = 0$.
On the other hand, integrating by parts once, we have: $\langle \nabla\, z_{\mathbf{f}'}, \tilde{\mathbf{v}}_q \rangle_{\mathbf{H}^1_0(\Omega)} = -\,(z_{\mathbf{f}'}, q)_{L^2(\Omega)}$, so the claim follows. ◇

As a consequence of (7.7.6), we have the

Corollary 7.7.1. *The right-hand-side ℓ_λ is equal to $(\mathbf{v}, q) \mapsto \lambda \langle \mathbf{f}, \mathbf{v} \rangle_{\mathbf{H}^{-1}(\Omega), \mathbf{H}^1_0(\Omega)} + \nu^{-1}(z_{\mathbf{f}}, q)_{L^2(\Omega)}$.*

7.7.1 New variational formulation using orthogonality

Going back to the original Problem $(7.2.1)_H$, we note that all solutions $(\mathbf{u}', p')$ to Problem (7.7.3) are such that $\mathbf{u}' \in \mathbf{V}$. So, in the statement of Problem (7.7.3), the term $(\mathbf{u}', \tilde{\mathbf{v}}_q)_{\mathbf{H}^1_0(\Omega)}$ can be removed by orthogonality from the expression (7.7.2) of the bilinear form $\tilde{a}_\lambda$: as a matter of fact, according to (4.3.1), for all $q \in L^2_{zmv}(\Omega)$, $(\mathbf{u}', \tilde{\mathbf{v}}_q)_{\mathbf{H}^1_0(\Omega)} = 0$ because $\tilde{\mathbf{v}}_q \in \mathbf{V}^\perp$. Interestingly, in this manner one stabilizes the expression (7.2.6) in the proof of Proposition 7.2.2, because the cross term vanishes there: no treatment is required and the resulting bilinear form is coercive for all $\lambda > 0$. So, we propose to remove the second term in the expression (7.7.2) of the bilinear form with operator T_λ, that is, we introduce the bilinear form $a_\lambda(\cdot,\cdot)$ such that:

$$\left\{ \begin{array}{rcl} a_\lambda : \quad \mathcal{X} \times \mathcal{X} & \to & \mathbb{R} \\ ((\mathbf{u}', p'), (\mathbf{v}, q)) & \mapsto & \nu\, \lambda\, (\mathbf{u}', \mathbf{v})_{\mathbf{H}^1_0(\Omega)} + \nu^{-1}\, (p', q)_{L^2(\Omega)} \\ & & +\lambda \left[(q, \operatorname{div} \mathbf{u}')_{L^2(\Omega)} - (p', \operatorname{div} \mathbf{v})_{L^2(\Omega)} \right] \end{array} \right. . \tag{7.7.7}$$

Proposition 7.7.3. *The bilinear form* (7.7.7) *is coercive.*

Proof. We have: $a_\lambda\left((\mathbf{u}', p'), (\mathbf{u}', p')\right) = \nu\, \lambda\, \|\mathbf{u}'\|^2_{\mathbf{H}^1_0(\Omega)} + \nu^{-1}\, \|p'\|^2_{L^2(\Omega)} \geq \nu\, \min(1, \lambda)\, \|(\mathbf{u}', p')\|^2_{\mathcal{X},\nu}$. ◇

With the help of explicit T-coercivity and using orthogonality, we can now propose a second new variational formulation to Problem $(7.2.1)_H$, which reads:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{u}, p) \in \mathcal{X} \text{ such that} \\ a_\lambda\left((\mathbf{u}, p), (\mathbf{v}, q)\right) = \lambda \langle \mathbf{f}, \mathbf{v} \rangle_{\mathbf{H}^1_0(\Omega)} + \nu^{-1}\, (z_\mathbf{f}, q)_{L^2(\Omega)} \quad \forall (\mathbf{v}, q) \in \mathcal{X}. \end{array} \right. \tag{7.7.8}$$

By contrast with remark 7.2.1, the result below holds for all $\lambda > 0$.

Theorem 7.7.1. *For all $\lambda > 0$, Problem* (7.7.8) *is well-posed and is equivalent to Problem* $(7.2.1)_H$ *with* $g = 0$.

Proof. The bilinear form $a_\lambda(\cdot,\cdot)$ is continuous and coercive. Let $\mathbf{f} \in \mathbf{H}^{-1}(\Omega)$, and let $z_\mathbf{f}$ be given by (7.7.5), the linear form $\ell_\lambda(\cdot)$ is continuous. According to Lax-Milgram Theorem, Problem (7.7.8) is well-posed. It exists a unique solution $(\mathbf{u}, p)$ which depends continuously on the data. Regarding the equivalence with Problem $(7.2.1)_H$ with $g = 0$, we already observed that solving (7.7.3) is equivalent to solving (7.2.4), and that both of them are equivalent to solving Problem $(7.2.1)_H$ with $g = 0$. Then, if $(\mathbf{u}, p)$ solves Problem $(7.2.1)_H$ with data $\mathbf{f} \in \mathbf{H}^{-1}(\Omega)$ and $g = 0$, one has in particular that $\mathbf{u} \in \mathbf{V}$. Hence it follows from the above that $(\mathbf{u}, p)$ solves (7.7.8) with data $\mathbf{f}$ and $g = 0$, resp. $z_\mathbf{f}$ given by (7.7.5). Conversely, assume that $(\mathbf{u}, p)$ solves (7.7.8), with data $\mathbf{f} \in \mathbf{H}^{-1}(\Omega)$ and $z_\mathbf{f}$ given by (7.7.5). Denote by $(\mathbf{u}^\dagger, p^\dagger)$ the solution to Problem $(7.2.1)_H$ with data $\mathbf{f}$. According to the above and by linearity, $(\mathbf{u} - \mathbf{u}^\dagger, p - p^\dagger)$ solves (7.7.8) with vanishing data, hence to is equal to $(\mathbf{0}, 0)$ by uniqueness: in other words, $(\mathbf{u}, p)$ solves Problem $(7.2.1)_H$ with data $\mathbf{f}$. ◇

Notice that we can write problem (7.7.8) with two equations as:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{u}, p) \in \mathcal{X} \text{ s.t. for all } (\mathbf{v}, q) \in \mathcal{X} \\ (i) \quad \nu\, \lambda (\mathbf{u}, \mathbf{v})_{\mathbf{H}^1_0(\Omega)} - \lambda (p, \operatorname{div} \mathbf{v})_{L^2(\Omega)} = \lambda \langle \mathbf{f}, \mathbf{v} \rangle_{\mathbf{H}^1_0(\Omega)}, \\ (ii)\ \lambda (q, \operatorname{div} \mathbf{u})_{L^2(\Omega)} + \nu^{-1}\, (p, q)_{L^2(\Omega)} = \nu^{-1}\, (z_\mathbf{f}, q)_{L^2(\Omega)}. \end{array} \right. \tag{7.7.9}$$

This new variational formulation appears as a stabilized variational formulation, in the sense of §II.1.2 in [161], pages 120-123. It can be solved once $z_\mathbf{f}$, or a suitable approximation of $z_\mathbf{f}$, is available. This suggests to use (7.7.9) as a *post processing step* as follows:

- Compute $z_\mathbf{f}$ by solving numerically the Stokes problem (7.2.2) with $\nu = 1$, and data $\mathbf{f}$;
- Solve (7.7.9) with the data $\mathbf{f}$ and the computed $z_\mathbf{f}$.

Notice that we can recover that the solution $\mathbf{u}$ to (7.7.9) belongs to $\mathbf{V}$ in a simple manner: let us split $\mathbf{u} = \mathbf{u}_0 + \mathbf{u}_\perp$, where $(\mathbf{u}_0, \mathbf{u}_\perp) \in \mathbf{V} \times \mathbf{V}^\perp$, so that $\operatorname{div} \mathbf{u} = \operatorname{div} \mathbf{u}_\perp$. Choosing $\mathbf{v} = \mathbf{u}_\perp$ in (7.7.9)-(i), and

$q = \operatorname{div} \mathbf{u}$ in (7.7.9)-(ii), we obtain:

$$\left\{ \begin{array}{ll} \nu \left\| \mathbf{u}_\perp \right\|^2_{\mathbf{H}^1_0(\Omega)} - (p, \operatorname{div} \mathbf{u})_{L^2(\Omega)} = \langle \mathbf{f}, \mathbf{u}_\perp \rangle_{\mathbf{H}^1_0(\Omega)} = -(z_{\mathbf{f}}, \operatorname{div} \mathbf{u})_{L^2(\Omega)} & \text{(i)} \\ \lambda \| \operatorname{div} \mathbf{u} \|^2_{L^2(\Omega)} + \nu^{-1} (p, \operatorname{div} \mathbf{u})_{L^2(\Omega)} = \nu^{-1} (z_{\mathbf{f}}, \operatorname{div} \mathbf{u})_{L^2(\Omega)} & \text{(ii)} \end{array} \right. . \tag{7.7.10}$$

Summing (7.7.10)-(i) and (7.7.10)-(ii) times ν, we obtain: $\nu \left(\left\| \mathbf{u}_\perp \right\|^2_{\mathbf{H}^1_0(\Omega)} + \lambda \| \operatorname{div} \mathbf{u} \|^2_{L^2(\Omega)} \right) = 0$. Hence, $\mathbf{u}_\perp = 0$ and $\operatorname{div} \mathbf{u} = 0$.

7.7.2 Numerical algorithms

We solve the classical incompressible Stokes model, with $g = 0$. We propose below two strategies depending on whether $z_{\mathbf{f}}$ is known or not. Below, we consider homogeneous Dirichlet boundary conditions (see appendix B.4 for nonhomogeneous Dirichlet boundary conditions).
If $z_{\mathbf{f}}$ is known, one solves (7.7.9) directly. If this is not the case, we propose to start from some approximate value of $z_{\mathbf{f}}$, and then to iterate by solving a series of problems like (7.7.9).
So, after discretization, if $z_{\mathbf{f}}$ is known, one solves a linear system like

$$\left\{ \begin{array}{l} \text{Find } (\underline{U}, \underline{P}) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p} \text{ such that:} \\ \begin{array}{rcl} \nu\lambda \mathbb{A}\, \underline{U} - \lambda \mathbb{B}^T \underline{P} & = & \lambda \underline{F} \\ \lambda \mathbb{B}\, \underline{U} + \nu^{-1} \mathbb{M}\, \underline{P} & = & \nu^{-1} \mathbb{M}\, \underline{Z} \end{array} \end{array} \right. \tag{7.7.11}$$

where $\underline{U} \in \mathbb{R}^{N_u}$ represents the discrete velocity, resp. $\underline{P} \in \mathbb{R}^{N_p}$ the discrete pressure, while $\underline{F} \in \mathbb{R}^{N_u}$ stands for $\mathbf{f}$, and $\underline{Z} \in \mathbb{R}^{N_p}$ stands for $z_{\mathbf{f}}$. Classically, $\underline{U} \mapsto (\mathbb{A}\underline{U} | \underline{U})$ measures discrete velocities, $\underline{P} \mapsto (\mathbb{M}\underline{P} | \underline{P})$ measures discrete pressures and, as a consequence, $\underline{U} \mapsto (\mathbb{B}\underline{U} | \mathbb{M}^{-1}\mathbb{B}\underline{U})$ measures the divergence of the discrete velocities.
On the other hand, if $z_{\mathbf{f}}$ is not known, starting from an initial guess $\underline{P}^{-1} \in \mathbb{R}^{N_p}$, for $n = 0, 1, \ldots$, one solves linear systems like

$$\left\{ \begin{array}{l} \text{Find } (\underline{U}^n, \underline{P}^n) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p} \text{ such that:} \\ \begin{array}{rcl} \nu\lambda \mathbb{A}\, \underline{U}^n - \lambda \mathbb{B}^T \underline{P}^n & = & \lambda \underline{F}_u \\ \lambda \mathbb{B}\, \underline{U}^n + \nu^{-1} \mathbb{M}\, \underline{P}^n & = & \nu^{-1} \mathbb{M}\, \underline{P}^{n-1} \end{array} \end{array} \right. . \tag{7.7.12}$$

In this situation, one needs to prove that the iterative solver is actually converging, and to provide a stopping criterion.

Remark 7.7.1. *The matrix of Algorithm* (7.7.12) *is similar to the one of the first-order artificial compressibility algorithm [232, 169, 170]. In [169] and [170], the space discretization is done with the MAC approximation on a Cartesian grid [172]. In [170], it is also done with the Taylor-Hood finite element [239].*

Theorem 7.7.2 (`Convergence`)**.** *Assume that the matrices $\mathbb{A} \in \mathbb{R}^{N_u \times N_u}$ and $\mathbb{M} \in \mathbb{R}^{N_p \times N_p}$ are symmetric positive-definite. Then, the sequence $(\underline{U}^n, \underline{P}^n)_n$ of solutions to* (7.7.12) *is converging, and it converges to $(\underline{U}^\infty, \underline{P}^\infty) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p}$ which is governed by*

$$\left\{ \begin{array}{l} \textit{Find } (\underline{U}^\infty, \underline{P}^\infty) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p} \textit{ such that:} \\ \begin{array}{rcl} \nu \mathbb{A}\, \underline{U}^\infty - \mathbb{B}^T \underline{P}^\infty & = & \underline{F}_u \\ \mathbb{B}\, \underline{U}^\infty & = & 0 \\ \mathbb{M}\, (\underline{P}^\infty - \underline{P}^{-1}) \in \operatorname{im}(\mathbb{B}) & & \end{array} \end{array} \right. . \tag{7.7.13}$$

Proof. In the proof, we use the change of variable $\underline{Y} := \mathbb{M}^{\frac{1}{2}} \underline{P} \in \mathbb{R}^{N_p}$, and specifically $\underline{Y}^n = \mathbb{M}^{\frac{1}{2}} \underline{P}^n$ for $n \geq -1$. Then let $\delta \underline{P}^n = \underline{P}^n - \underline{P}^{n-1}$ and $\delta \underline{Y}^n := (\mathbb{M}^{\frac{1}{2}} \delta \underline{P}^n)$ for $n \geq 0$.
1/ We notice first that the following recursive relation holds:

$$\forall n \geq 1, \quad (\mathbb{M} + \lambda \mathbb{B} \mathbb{A}^{-1} \mathbb{B}^T) \delta \underline{P}^n = \mathbb{M} \delta \underline{P}^{n-1}, \tag{7.7.14}$$

where the matrix $\lambda \mathbb{B}\mathbb{A}^{-1}\mathbb{B}^T$ is a symmetric positive matrix. By construction, $\delta \underline{Y}^n = -\nu\lambda \mathbb{M}^{-\frac{1}{2}}\mathbb{B}\,\underline{U}^n \in \mathrm{im}(\mathbb{M}^{-\frac{1}{2}}\mathbb{B})$. Since $\mathbb{R}^{N_p} = \mathrm{im}(\mathbb{M}^{-\frac{1}{2}}\mathbb{B}) \overset{\perp}{\oplus} \ker(\mathbb{B}^T\mathbb{M}^{-\frac{1}{2}})$, one has $\delta \underline{Y}^n \in \ker(\mathbb{B}^T\mathbb{M}^{-\frac{1}{2}})^\perp$ and (7.7.14) reads:

$$\forall n \geq 1, \quad (\mathbb{I} + \lambda \mathbb{M}^{-\frac{1}{2}}\mathbb{B}\mathbb{A}^{-1}\mathbb{B}^T\mathbb{M}^{-\frac{1}{2}})\delta \underline{Y}^n = \delta \underline{Y}^{n-1}. \tag{7.7.15}$$

The iteration matrix $(\mathbb{I} + \lambda \mathbb{M}^{-\frac{1}{2}}\mathbb{B}\mathbb{A}^{-1}\mathbb{B}^T\mathbb{M}^{-\frac{1}{2}})^{-1}$ restricted to $\ker(\mathbb{B}^T\mathbb{M}^{-\frac{1}{2}})^\perp$ is a symmetric positive definite matrix, whose eigenvalues are all strictly smaller than 1. Hence, its spectral radius ρ_λ is strictly smaller than 1.
2/ We deduce that it exists $C_\lambda \in (0,1)$ independent of the data and of n such that for all $n \geq 1$, $|\delta \underline{Y}^n|_2 \leq C_\lambda |\delta \underline{Y}^{n-1}|_2$. Thus, the series $\sum_{n\geq 1} \delta \underline{Y}^n$ is a convergent series. Noticing that $\sum_{n=0}^N \delta \underline{Y}^n = \underline{Y}^N - \underline{Y}^{-1}$, we conclude that the sequence $(\underline{Y}^n)_n$ is a convergent sequence and we set $\underline{Y}^\infty = \lim_{n\to\infty} \underline{Y}^n$. By construction $\underline{Y}^\infty - \underline{Y}^{-1} \in \mathrm{im}(\mathbb{M}^{-\frac{1}{2}}\mathbb{B}) = \ker(\mathbb{B}^T\mathbb{M}^{-\frac{1}{2}})^\perp$, and for all n, the algorithm doesn't change the orthogonal projection of $\underline{Y}^n$ onto $\ker(\mathbb{B}^T\mathbb{M}^{-\frac{1}{2}})$, which is equal to that of $\underline{Y}^{-1}$ onto $\ker(\mathbb{B}^T\mathbb{M}^{-\frac{1}{2}})$.
3/ Because $\mathbb{M}^{\frac{1}{2}}$ is invertible, we obtain that the sequence $(\underline{P}^n)_n$ is also a convergent sequence and we call $\underline{P}^\infty = \mathbb{M}^{-\frac{1}{2}}\underline{Y}^\infty$ its limit. In the same way, noting that $\mathbb{A}$ is invertible, we obtain that the sequence $(\underline{U}^n)_n$ is a convergent sequence and we call $\underline{U}^\infty = \nu^{-1}\mathbb{A}^{-1}(\underline{F}_u + \mathbb{B}^T\underline{P}^\infty)$ its limit.
Passing to the limit, it holds that $\mathbb{B}\,\underline{U}^\infty = 0$ and $\mathbb{M}(\underline{P}^\infty - \underline{P}^{-1}) \in \mathrm{im}(\mathbb{B})$. ⋄

Remark 7.7.2. *In the above proof, we notice that the mapping $\lambda \mapsto \rho_\lambda$ is monotonically decreasing. In §7.7.5 dedicated to the numerical experiments, we set the value of λ equal to 1 or 10.*

Remark 7.7.3. *There are two exclusive cases for Problem* (7.7.13)*:*

- *either,* $\mathrm{im}(\mathbb{B}) = \mathbb{R}^{N_p}$*: a discrete inf-sup condition holds. Uniqueness of $\underline{P}^\infty$ is guaranteed and the last line is trivial;*

- *or,* $\mathrm{im}(\mathbb{B}) \subsetneq \mathbb{R}^{N_p}$*: there is no discrete inf-sup condition. However in this case, the last line guarantees the uniqueness of $\underline{P}^\infty$. As a matter of fact, the algorithm does not change the orthogonal projection of $\underline{P}^n$ onto* $\ker(\mathbb{B}^T)$*, which is equal to that of $\underline{P}^{-1}$ onto* $\ker(\mathbb{B}^T)$*, where orthogonality is understood with respect to the inner product* $(\mathbb{M}\cdot|\cdot)$*.*

Let us consider that the assumptions of Theorem 7.7.2 hold. A critical ingredient is to define the stopping criterion, in particular what are the relevant quantities of interest. By definition, $(\mathbb{M}\delta \underline{P}^n|\delta \underline{P}^n) = |\delta \underline{Y}^n|_2^2$ is strictly monotonically decreasing. Hence, denoting by $|\cdot|_{\mathbb{A}} : \underline{U} \mapsto (\mathbb{A}\underline{U}|\underline{U})^{1/2}$ the norm of the discrete velocity, resp. $|\cdot|_{\mathbb{M}} : \underline{P} \mapsto (\mathbb{M}\underline{P}|\underline{P})^{1/2}$ the norm of the discrete pressure, we define the stopping criterion by comparing the adimensionalized quantities $|\delta \underline{P}^n|_{\mathbb{M}}$ and $|\underline{U}^n|_{\mathbb{A}}$ (we note that both quantities are easily computable). This amounts to setting the stopping criterion as

$$|\delta \underline{P}^n|_{\mathbb{M}} \leq \epsilon\, |\underline{U}^n|_{\mathbb{A}} \tag{7.7.16}$$

for ad hoc $\epsilon > 0$. In our numerical experiments, we fixed $n = 1$ or $n = 8$ with $\epsilon = 10^{-12}$.
Finally, we recall that $\mathbb{B}\underline{U}^n$ stands for the divergence of the discrete velocity $\mathrm{div}\,\mathbf{u}_h^n$, and that one aims at diminishing the value of its norm. Considering again the auxiliary variable $\delta \underline{Y}^n$ introduced in the proof of Theorem 7.7.2, we find:

$$\lambda^2 |\mathbb{M}^{-1/2}\mathbb{B}\underline{U}^n|_2^2 = \nu^{-2}|\delta \underline{Y}^n|_2^2.$$

As a consequence of the proof of the previous Theorem, we infer that the sequence of norms $(|\mathbb{M}^{-1/2}\mathbb{B}\underline{U}^n|_2)_n$ is strictly monotonically decreasing, and so is $(\|\mathrm{div}\,\mathbf{u}_h^n\|_{L^2(\Omega)})_n$.

7.7.3 Discretization

We solve numerically the new variational formulation, written like in Problem (7.7.9). Below, we consider homogeneous Dirichlet boundary conditions (see appendix B.4 for nonhomogeneous Dirichlet boundary conditions).

Discretizations

Consider $(\mathcal{T}_h)_h$ a simplicial triangulation sequence of Ω. We use the notations of Section 4.7. We define the conforming spaces of P^k-Lagrange functions:

$$V_h^k := \left\{ v_h \in H^1(\Omega); \quad \forall T \in \mathcal{T}_h,\, v_{h|T} \in P^k(T) \right\}, \quad V_{0,h}^k := \left\{ v_h \in V_h^k; \quad v_{h|\partial\Omega} = 0 \right\}.$$

We set $\mathbf{V}_h^k := (V_h^k)^d$ and $\mathbf{V}_{0,h}^k := (V_{0,h}^k)^d$. We call $Q_h^k := P_{disc}^k(\mathcal{T}_h) \cap L^2_{zmv}(\Omega)$. Notice that $\mathbf{V}_{0,h}^k \times Q_h^{k'} \subset \mathbf{H}_0^1(\Omega) \times L^2_{zmv}(\Omega)$ and $\operatorname{div} \mathbf{V}_{0,h}^k \subset Q_h^{k-1}$.

Let $k \geq 1$. The (conforming) discretization of Problem (7.7.9) with Taylor-Hood $\mathbf{P}^k \times P_{disc}^{k-1}$ finite element reads:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{u}_h, p_h) \in \mathbf{V}_{0,h}^k \times Q_h^{k-1} \text{ s.t. for all } (\mathbf{v}_h, q_h) \in \mathbf{V}_{0,h}^k \times Q_h^{k-1} : \\ \nu\, \lambda(\mathbf{u}_h, \mathbf{v}_h)_{\mathbf{H}_0^1(\Omega)} - \lambda(p_h, \operatorname{div} \mathbf{v}_h)_{L^2(\Omega)} = \lambda \langle \mathbf{f}, \mathbf{v}_h \rangle_{\mathbf{H}_0^1(\Omega)}, \\ \lambda(q_h, \operatorname{div} \mathbf{u}_h)_{L^2(\Omega)} + \nu^{-1} (p_h, q_h)_{L^2(\Omega)} = \nu^{-1} (z_{\mathbf{f}}, q_h)_{L^2(\Omega)}. \end{array} \right. \tag{7.7.17}$$

Write $\mathbf{f} = -\nu \Delta \mathbf{u} + \nabla z_{\mathbf{f}}$ (recall (7.2.1) $z_{\mathbf{f}} = p$), so that Problem (7.7.17) reads:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{u}_h, p_h) \in \mathbf{V}_{0,h}^k \times Q_h^{k-1} \text{ s.t. for all } (\mathbf{v}_h, q_h) \in \mathbf{V}_{0,h}^k \times Q_h^{k-1} : \\ \nu\, \lambda(\mathbf{u}_h, \mathbf{v}_h)_{\mathbf{H}_0^1(\Omega)} - \lambda(p_h, \operatorname{div} \mathbf{v}_h)_{L^2(\Omega)} = \nu\, \lambda(\mathbf{u}, \mathbf{v}_h)_{\mathbf{H}_0^1(\Omega)} - \lambda(z_{\mathbf{f}}, \operatorname{div} \mathbf{v}_h)_{L^2(\Omega)}, \\ \lambda(q_h, \operatorname{div} \mathbf{u}_h)_{L^2(\Omega)} + \nu^{-1} (p_h, q_h)_{L^2(\Omega)} = \nu^{-1} (z_{\mathbf{f}}, q_h)_{L^2(\Omega)}. \end{array} \right. \tag{7.7.18}$$

Remark 7.7.4. *It is well known that discretizing* (7.2.2) *with the Taylor-Hood $\mathbf{P}^k \times P_{disc}^{k-1}$ finite element, $k \geq 1$, is not stable on all shape regular meshes [50]. A wide range of strategies to get round this problem have been explored for years. Below is a quick overview of these strategies:*

- *The discrete velocity space can be enriched see [44, 15, 24, 83, 193] for $k = 1$.*
- *The discrete variational formulation can be stabilized, see [47, 25, 249, 5] for $k = 1$, [3] for $k = 2$.*
- *The mesh can be designed in such a way that the uniform discrete inf-sup condition applies, see [149] for $k = 1$, [230, 221, 23] for $k > 1$.*

Error estimate when $z_{\mathbf{f}}$ is known

Let $\Pi_{h,cg}$ be the $\mathbf{L}^2(\Omega)$-orthogonal projection operator from $\mathbf{L}^2(\Omega)$ to $\mathbf{V}_{0,h}^k$ and $\Pi_{h,dg}$ be the $L^2(\Omega)$-orthogonal projection operator from $L^2(\Omega)$ to $\mathcal{P}_{disc}^{k-1}(\mathcal{T}_h)$. We have the

Proposition 7.7.4. *Let $(\mathbf{u}, p) \in \mathcal{X}$ be the solution to Problem* (7.7.9) *and $(\mathbf{u}_h, p_h) \in \mathcal{X}_h$ be the solution to Problem* (7.7.17)*. We have the following error estimate:*

$$\|(\mathbf{u}_h - \mathbf{u}, p_h - \Pi_{h,dg} z_{\mathbf{f}})\|_{\mathcal{X}, \sqrt{\lambda}\nu} \leq (1 + \sqrt{d\lambda})\, \|\mathbf{u} - \Pi_{h,cg}\mathbf{u}\|_{\mathbf{H}_0^1(\Omega)}. \tag{7.7.19}$$

Suppose that $\mathbf{u} \in \mathbf{P}^k(\Omega)$. Then $\mathbf{u}_h = \mathbf{u}$ and $p_h = \Pi_{h,dg} z_{\mathbf{f}}$.

Notice that the error estimate for the velocity is fully independent of the pressure. Hence, the (conforming) discretization of Problem (7.7.9) with $\mathbf{P}^k \times P_{disc}^{k-1}$ and for which $z_{\mathbf{f}}$ is known is a so called *pressure robust method* for which the discrete velocity $\mathbf{u}_h$ tends to $\mathbf{u}$ independently of ν. The discrete pressure p_h tends to $\Pi_{h,dg} z_{\mathbf{f}}$ all the faster the smaller ν is. Interestingly, we remark that if $\mathbf{u} = 0$, then $\Pi_{h,cg}\mathbf{u} = 0$, so that $\mathbf{u}_h = 0$ and $p_h = \Pi_{h,dg} z_{\mathbf{f}}$, even for $k = 1$ (see Section 7.7.5).
Let us now prove Proposition 7.7.4.

Proof. Setting $z_{\mathbf{f},h} = \Pi_{h,dg} z_{\mathbf{f}}$, we have: $(z_{\mathbf{f}}, \operatorname{div} \mathbf{v}_h)_{L^2(\Omega)} = (z_{\mathbf{f},h}, \operatorname{div} \mathbf{v}_h)_{L^2(\Omega)}$ and $(z_{\mathbf{f}}, q_h)_{L^2(\Omega)} = (z_{\mathbf{f},h}, q_h)_{L^2(\Omega)}$ for all $\mathbf{v}_h \in \mathbf{V}_{0,h}^k$, and all $q_h \in Q_h^{k-1}$. Summing both equations of Problem (7.7.18) and reshuffling terms, it comes:

$$\left\{ \begin{array}{l} \text{Find } (\mathbf{u}_h, p_h) \in \mathbf{V}_{0,h}^k \times Q_h^{k-1} \text{ s.t. for all } (\mathbf{v}_h, q_h) \in \mathbf{V}_{0,h}^k \times Q_h^{k-1} : \\ \nu\, \lambda(\mathbf{u}_h - \mathbf{u}, \mathbf{v}_h)_{\mathbf{H}_0^1(\Omega)} + \nu^{-1} (p_h - z_{\mathbf{f},h}, q_h)_{L^2(\Omega)} = \lambda(p_h - z_{\mathbf{f},h}, \operatorname{div} \mathbf{v}_h)_{L^2(\Omega)} - \lambda(q_h, \operatorname{div} \mathbf{u}_h)_{L^2(\Omega)}. \end{array} \right. \tag{7.7.20}$$

Choosing $\mathbf{v}_h = \mathbf{u}_h - \Pi_{h,cg}\mathbf{u} = (\mathbf{u}_h - \mathbf{u}) + (\mathbf{u} - \Pi_{h,cg}\mathbf{u})$ and $q_h = p_h - z_{\mathbf{f},h}$ in (7.7.20), and dividing by $\lambda\,\nu$, we obtain:

$$\|\mathbf{u}_h - \mathbf{u}\|^2_{\mathbf{H}^1_0(\Omega)} + (\,\mathbf{u}_h - \mathbf{u}, \mathbf{u} - \Pi_{h,cg}\mathbf{u}\,)_{\mathbf{H}^1_0(\Omega)} + \lambda^{-1}\nu^{-2}\|p_h - z_{\mathbf{f},h}\|^2_{L^2(\Omega)} = -\nu^{-1}(p_h - z_{\mathbf{f},h}, \operatorname{div}\Pi_{h,cg}\mathbf{u})_{L^2(\Omega)}.$$

Hence, using Cauchy-Schwarz inequality, letting $\delta\mathbf{u} = \mathbf{u}_h - \mathbf{u}$ we get:

$$\begin{array}{l}\|\delta\mathbf{u}\|^2_{\mathbf{H}^1_0(\Omega)} + \lambda^{-1}\nu^{-2}\|p_h - z_{\mathbf{f},h}\|^2_{L^2(\Omega)}\\ \le \|\delta\mathbf{u}\|_{\mathbf{H}^1_0(\Omega)}\,\|\mathbf{u} - \Pi_{h,cg}\mathbf{u}\|_{\mathbf{H}^1_0(\Omega)} + \nu^{-1}\|p_h - z_{\mathbf{f},h}\|_{L^2(\Omega)}\,\|\operatorname{div}\Pi_{h,cg}\mathbf{u}\|_{L^2(\Omega)}\\ \le (\|\delta\mathbf{u}\|^2_{\mathbf{H}^1_0(\Omega)} + \lambda^{-1}\nu^{-2}\|p_h - z_{\mathbf{f},h}\|^2_{L^2(\Omega)})^{1/2}(\|\mathbf{u} - \Pi_{h,cg}\mathbf{u}\|^2_{\mathbf{H}^1_0(\Omega)} + \lambda\|\operatorname{div}\Pi_{h,cg}\mathbf{u}\|^2_{L^2(\Omega)})^{1/2}.\end{array}$$

Finally $\|\operatorname{div}\Pi_{h,cg}\mathbf{u}\|_{L^2(\Omega)} \le \sqrt{d}\,\|\mathbf{u} - \Pi_{h,cg}\mathbf{u}\|_{\mathbf{H}^1_0(\Omega)}$, so we have: $\|(\mathbf{u}_h - \mathbf{u}, p_h - z_{\mathbf{f},h})\|_{\mathcal{X},\sqrt{\lambda}\nu} \le \sqrt{1+d\lambda}\,\|\mathbf{u} - \Pi_{h,cg}\mathbf{u}\|_{\mathbf{H}^1_0(\Omega)}$.
If $\mathbf{u} \in \mathbf{P}^k(\Omega)$, then $\Pi_{h,cg}\mathbf{u} = \mathbf{u}$ and the right-hand side is equal to 0. Hence, $\mathbf{u}_h = \mathbf{u}$ and $p_h = z_{\mathbf{f},h}$. ⋄

Error estimate when $z_{\mathbf{f}}$ is not known

If $z_{\mathbf{f}}$ is not known explicitly, let us assume that we have at hand an approximation $p_h^{old} \in Q_h^{k-1}$ and use it to solve the following approximation of Problem (7.7.9):

$$\left\{\begin{array}{l}\text{Find } (\tilde{\mathbf{u}}, \tilde{p}) \in \mathcal{X} \text{ s.t. for all } (\mathbf{v}, q) \in \mathcal{X}:\\ \nu\,\lambda(\tilde{\mathbf{u}}, \mathbf{v})_{\mathbf{H}^1_0(\Omega)} - \lambda(\tilde{p}, \operatorname{div}\mathbf{v})_{L^2(\Omega)} = \lambda\langle\mathbf{f}, \mathbf{v}\rangle_{\mathbf{H}^1_0(\Omega)},\\ \lambda(q, \operatorname{div}\tilde{\mathbf{u}})_{L^2(\Omega)} + \nu^{-1}\,(\tilde{p}, q)_{L^2(\Omega)} = \nu^{-1}\,(p_h^{old}, q)_{L^2(\Omega)}.\end{array}\right. \tag{7.7.21}$$

Let us write again $\mathbf{f} = -\nu\Delta\mathbf{u} + \nabla\, z_{\mathbf{f}}$, so that Problem (7.7.21) reads:

$$\left\{\begin{array}{l}\text{Find } (\tilde{\mathbf{u}}, \tilde{p}) \in \mathcal{X} \text{ s.t. for all } (\mathbf{v}, q) \in \mathcal{X}:\\ \nu\,\lambda(\tilde{\mathbf{u}} - \mathbf{u}, \mathbf{v})_{\mathbf{H}^1_0(\Omega)} - \lambda(\tilde{p} - z_{\mathbf{f}}, \operatorname{div}\mathbf{v})_{L^2(\Omega)} = 0,\\ \lambda(q, \operatorname{div}\tilde{\mathbf{u}})_{L^2(\Omega)} + \nu^{-1}\,(\tilde{p} - p_h^{old}, q)_{L^2(\Omega)} = 0.\end{array}\right. \tag{7.7.22}$$

Proposition 7.7.5. *Let $(\mathbf{u}, p) \in \mathcal{X}$ be the solution to Problem (7.7.9) and $(\tilde{\mathbf{u}}, \tilde{p}) \in \mathcal{X}$ be the solution to Problem (7.7.21). We have the following estimates:*

$$\left\{\begin{array}{lcl}\|\tilde{p} - z_{\mathbf{f}}\|_{L^2(\Omega)} & \le & \|z_{\mathbf{f}} - p_h^{old}\|_{L^2(\Omega)},\\ \|\tilde{\mathbf{u}} - \mathbf{u}\|_{\mathbf{H}^1_0(\Omega)} & \le & (\sqrt{\lambda}\,\nu)^{-1}\|z_{\mathbf{f}} - p_h^{old}\|_{L^2(\Omega)}.\end{array}\right. \tag{7.7.23}$$

Proof. Choosing $\mathbf{v} = \tilde{\mathbf{u}} - \mathbf{u}$ and $q = \tilde{p} - z_{\mathbf{f}}$ in (7.7.22), summing both equations and dividing by $\lambda\,\nu$, it comes:

$$\|\tilde{\mathbf{u}} - \mathbf{u}\|^2_{\mathbf{H}^1_0(\Omega)} + \lambda^{-1}\nu^{-2}\,(\tilde{p} - p_h^{old}, \tilde{p} - z_{\mathbf{f}})_{L^2(\Omega)} = 0.$$

Writing $\tilde{p} - p_h^{old} = (\tilde{p} - z_{\mathbf{f}}) + (z_{\mathbf{f}} - p_h^{old})$, we obtain:

$$\|(\tilde{\mathbf{u}} - \mathbf{u}, \tilde{p} - z_{\mathbf{f}})\|^2_{\mathcal{X},\sqrt{\lambda}\nu} = -\lambda^{-1}\nu^{-2}(z_{\mathbf{f}} - p_h^{old}, \tilde{p} - z_{\mathbf{f}})_{L^2(\Omega)} \le \lambda^{-1}\nu^{-2}\|z_{\mathbf{f}} - p_h^{old}\|_{L^2(\Omega)}\,\|\tilde{p} - z_{\mathbf{f}}\|_{L^2(\Omega)}.$$

We obtain successively estimates (7.7.23). ⋄

The discretization of Problem (7.7.21) with $\mathbf{P}^k - P^{k-1}_{disc}$ finite element reads:

$$\left\{\begin{array}{l}\text{Find } (\tilde{\mathbf{u}}_h, p_h^{new}) \in \mathbf{V}^k_{0,h} \times Q_h^{k-1} \text{ s.t. for all } (\mathbf{v}_h, q_h) \in \mathbf{V}^k_{0,h} \times Q_h^{k-1}:\\ \nu\,\lambda(\tilde{\mathbf{u}}_h, \mathbf{v}_h)_{\mathbf{H}^1_0(\Omega)} - \lambda(p_h^{new}, \operatorname{div}\mathbf{v}_h)_{L^2(\Omega)} = \lambda\langle\mathbf{f}, \mathbf{v}_h\rangle_{\mathbf{H}^1_0(\Omega)},\\ \lambda(q_h, \operatorname{div}\tilde{\mathbf{u}}_h)_{L^2(\Omega)} + \nu^{-1}\,(p_h^{new}, q_h)_{L^2(\Omega)} = \nu^{-1}\,(p_h^{old}, q_h)_{L^2(\Omega)}.\end{array}\right. \tag{7.7.24}$$

Theorem 7.7.3. *Let $(\mathbf{u}, p) \in \mathcal{X}$ be the solution to Problem (7.7.9) and $(\tilde{\mathbf{u}}_h, p_h^{new}) \in \mathcal{X}$ be the solution to Problem (7.7.24). We have the following error estimate:*

$$\begin{array}{lcl}\|(\tilde{\mathbf{u}}_h - \mathbf{u}, p_h^{new} - \Pi_{h,dg}z_{\mathbf{f}})\|_{\mathcal{X},\sqrt{\lambda}\nu} & \le & \sqrt{d\lambda + 1}\|\mathbf{u} - \Pi_{h,cg}\mathbf{u}\|_{\mathbf{H}^1_0(\Omega)}\\ & & +(\sqrt{\lambda}\nu)^{-1}\,\|\Pi_{h,dg}z_{\mathbf{f}} - p_h^{old}\|_{L^2(\Omega)}.\end{array} \tag{7.7.25}$$

Suppose that $\mathbf{u} \in (P^k(\Omega))^d$. *Then* $\Pi_{h,cg}\mathbf{u} = \mathbf{u}$ *and we obtain the estimate below:*

$$\|(\tilde{\mathbf{u}}_h - \mathbf{u}, p_h^{new} - \Pi_{h,dg} z_{\mathbf{f}})\|_{\mathcal{X},\sqrt{\lambda}\nu} \leq (\sqrt{\lambda}\nu)^{-1} \|\Pi_{h,dg} z_{\mathbf{f}} - p_h^{old}\|_{L^2(\Omega)}. \tag{7.7.26}$$

In particular, if p_h^{old} is a convergent approximation of $\Pi_{h,dg} z_{\mathbf{f}}$, the solution $(\tilde{\mathbf{u}}_h, p_h^{new})$ is also a good approximation of $(\mathbf{u}, \Pi_{h,dg} z_{\mathbf{f}})$. Interestingly, the above with $k = 1$ corresponds to the $\mathbf{P}^1 \times P^0$ finite element pair, which is known to be unstable for the discretization of the usual variational formulation (7.2.2) of the Stokes problem.

Let us prove Theorem 7.7.3.

Proof. Setting $z_{\mathbf{f},h} = \Pi_{h,dg} z_{\mathbf{f}}$, we have: $\langle \mathbf{f}, \mathbf{v}_h \rangle_{\mathbf{H}_0^1(\Omega)} = \nu\,(\mathbf{u}, \mathbf{v}_h)_{\mathbf{H}_0^1(\Omega)} - (z_{\mathbf{f},h}, \operatorname{div} \mathbf{v}_h)_{L^2(\Omega)}$ for all $\mathbf{v}_h \in \mathbf{V}_{0,h}^k$. Hence, Problem (7.7.24) can be written as:

$$\left\{ \begin{array}{l} \text{Find } (\tilde{\mathbf{u}}_h, p_h^{new}) \in \mathbf{V}_{0,h}^k \times Q_h^{k-1} \text{ s.t. for all } (\mathbf{v}_h, q_h) \in \mathbf{V}_{0,h}^k \times Q_h^{k-1} : \\ \nu\,\lambda(\tilde{\mathbf{u}}_h - \mathbf{u}, \mathbf{v}_h)_{\mathbf{H}_0^1(\Omega)} - \lambda(p_h^{new} - z_{\mathbf{f},h}, \operatorname{div} \mathbf{v}_h)_{L^2(\Omega)} = 0, \\ \lambda(q_h, \operatorname{div} \tilde{\mathbf{u}}_h)_{L^2(\Omega)} + \nu^{-1}\,(p_h^{new} - p_h^{old}, q_h)_{L^2(\Omega)} = 0. \end{array} \right. \tag{7.7.27}$$

Choosing $(\mathbf{v}_h, q_h) = (\tilde{\mathbf{u}}_h - \Pi_{h,cg}\mathbf{u}, p_h^{new} - z_{\mathbf{f},h})$ in Problem (7.7.27), we have:

$$\left\{ \begin{array}{l} \nu\,\lambda(\tilde{\mathbf{u}}_h - \mathbf{u}, \tilde{\mathbf{u}}_h - \Pi_{h,cg}\mathbf{u})_{\mathbf{H}_0^1(\Omega)} - \lambda(p_h^{new} - z_{\mathbf{f},h}, \operatorname{div}(\tilde{\mathbf{u}}_h - \Pi_{h,cg}\mathbf{u}))_{L^2(\Omega)} = 0, \\ \lambda(p_h^{new} - z_{\mathbf{f},h}, \operatorname{div} \tilde{\mathbf{u}}_h)_{L^2(\Omega)} + \nu^{-1}\,(p_h^{new} - p_h^{old}, p_h^{new} - z_{\mathbf{f},h})_{L^2(\Omega)} = 0. \end{array} \right.$$

Summing both equations and dividing by $\lambda\,\nu$, it now comes:

$$(\tilde{\mathbf{u}}_h - \mathbf{u}, \tilde{\mathbf{u}}_h - \Pi_{h,cg}\mathbf{u})_{\mathbf{H}_0^1(\Omega)} + \lambda^{-1}\nu^{-2}\,(p_h^{new} - p_h^{old}, p_h^{new} - z_{\mathbf{f},h})_{L^2(\Omega)} + \nu^{-1}(p_h^{new} - z_{\mathbf{f},h}, \operatorname{div} \Pi_{h,cg}\mathbf{u})_{L^2(\Omega)} = 0.$$

Noticing that $\tilde{\mathbf{u}}_h - \Pi_{h,cg}\mathbf{u} = (\tilde{\mathbf{u}}_h - \mathbf{u}) + (\mathbf{u} - \Pi_{h,cg}\mathbf{u})$ and $p_h^{new} - p_h^{old} = (p_h^{new} - z_{\mathbf{f},h}) + (z_{\mathbf{f},h} - p_h^{old})$, we get:

$$\begin{array}{c} \|(\tilde{\mathbf{u}}_h - \mathbf{u}, p_h^{new} - z_{\mathbf{f},h})\|^2_{\mathcal{X},\sqrt{\lambda}\nu} + (\tilde{\mathbf{u}}_h - \mathbf{u}, \mathbf{u} - \Pi_{h,cg}\mathbf{u})_{\mathbf{H}_0^1(\Omega)} \\ +\lambda^{-1}\nu^{-2}\,(z_{\mathbf{f},h} - p_h^{old}, p_h^{new} - z_{\mathbf{f},h})_{L^2(\Omega)} + \nu^{-1}\,(p_h^{new} - z_{\mathbf{f},h}, \operatorname{div} \Pi_{h,cg}\mathbf{u})_{L^2(\Omega)} = 0. \end{array}$$

Using Cauchy-Schwarz, we deduce that:

$$\begin{array}{c} \|(\tilde{\mathbf{u}}_h - \mathbf{u}, p_h^{new} - z_{\mathbf{f},h})\|^2_{\mathcal{X},\sqrt{\lambda}\nu} \leq \|\tilde{\mathbf{u}}_h - \mathbf{u}\|_{\mathbf{H}_0^1(\Omega)} \|\mathbf{u} - \Pi_{h,cg}\mathbf{u}\|_{\mathbf{H}_0^1(\Omega)} \\ +(\sqrt{\lambda}\nu)^{-1} \|p_h^{new} - z_{\mathbf{f},h}\|_{L^2(\Omega)} \left((\sqrt{\lambda}\nu)^{-1} \|z_{\mathbf{f},h} - p_h^{old}\|_{L^2(\Omega)} + \sqrt{\lambda}\|\operatorname{div} \Pi_{h,cg}\mathbf{u}\|_{L^2(\Omega)} \right). \end{array}$$

Using again Cauchy-Schwarz inequality as in the proof of Proposition 7.7.4, together with $\|\operatorname{div} \Pi_{h,cg}\mathbf{u}\|_{L^2(\Omega)} \leq \sqrt{d}\,\|\mathbf{u} - \Pi_{h,cg}\mathbf{u}\|_{\mathbf{H}_0^1(\Omega)}$, we obtain the estimates (7.7.25) and (7.7.26). ◇

7.7.4 Discussion on the error indicators

As we noted after stating Theorem 7.7.3, if the error indicator on the pressure $\|z_{\mathbf{f},h} - p_h^{old}\|_{L^2(\Omega)}$ is small, so is $\|z_{\mathbf{f},h} - p_h^{new}\|_{L^2(\Omega)}$. Indeed, assuming that the errors on the velocities are negligible compared to these indicators, one derives from (7.7.25) the bound $\|z_{\mathbf{f},h} - p_h^{new}\|_{L^2(\Omega)} \leq \|z_{\mathbf{f},h} - p_h^{old}\|_{L^2(\Omega)}$. Interestingly, this situation is likely to occur when the viscosity is small, since there is a multiplicative factor equal to $(\sqrt{\lambda}\nu)^{-1}$ in front of the error indicators on the pressure. This further justifies the use of the iterative algorithm presented in Section 7.7.2 in this case.

7.7.5 Numerical illustration

In this Section, the computations are performed on a Dell Precision 3581 Intel Core i7 laptop. We propose some numerical experiments with $g = 0$, depending on whether or not $z_{\mathbf{f}}$ is known explicitly. In the latter case, we first compute some approximation p_h^{old}. In principle, either a classical, conforming or nonconforming, discretization can be used to compute p_h^{old}. Hence, we apply an iterative approach with a nonconforming initial computation followed by (several iterations of) our new formulation. we call the second part the post-processing. For each test-case, $(\mathbf{u}, p)$ is given, whereas the viscosity ν can vary (see the discussion after (7.2.3)). The numerical results are obtained on a github platform, implemented in

Octave language, see [G2].

Resolution algorithm

Consider nonconforming discretization of the classical variational formulation (7.2.2), cf. §7.4. Let $(\psi_i)_{i=1}^{N_u}$ be a basis of $\mathbf{X}_{0,h}$, and $(\phi_i)_{i=1}^{N_p}$ be a basis of Q_h. We set: $\underline{U} := (\underline{U}_i)_{i=1}^{N_u}$ where $\mathbf{u}_h := \sum_{i=1}^{N_u} \underline{U}_i \psi_i$, $\underline{P} := (\underline{P}_i)_{i=1}^{N_p}$ where $p_h := \sum_{i=1}^{N_p} \underline{P}_i \phi_i$, and $\underline{F}_u := (\underline{F}_{u,i})_{i=1}^{N_u}$. Then $\underline{F}_{u,i} = \ell_{\mathbf{f}}((\psi_i, 0))$ where $\ell_{\mathbf{f}}(\cdot)$ is defined in (7.4.9). Let $\mathbb{A} \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_u}$ be the velocity stiffness matrix, $\mathbb{B} \in \mathbb{R}^{N_p} \times \mathbb{R}^{N_u}$ be the velocity-pressure coupling matrix and $\mathbb{M} \in \mathbb{R}^{N_p} \times \mathbb{R}^{N_p}$ be the pressure mass matrix. The linear system to be solved is

$$\left\{ \begin{array}{l} \text{Find } (\underline{U}, \underline{P}) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p} \text{ such that:} \\ \nu \mathbb{A}\, \underline{U} - \mathbb{B}^T \underline{P} = \underline{F}_u \\ \qquad\qquad \mathbb{B}\, \underline{U} = 0 \end{array} \right. . \tag{7.7.28}$$

Let us set $\mathbb{K} = \mathbb{B}\, \mathbb{A}^{-1} \mathbb{B}^T \in \mathbb{R}^{N_p} \times \mathbb{R}^{N_p}$. The matrix $\mathbb{K}$ is a symmetric matrix, furthermore it is positive definite as soon as the kernel of $\mathbb{B}^T$ is reduced to $\{0\}$. When it is the case, the coupled velocity-pressure problem (7.7.28) can be solved using the three + one steps below (the fourth step being straightforward):

$$\begin{array}{ll} \text{Prediction:} & \text{Solve in } \underline{U}_\star \text{ such that } \nu\, \mathbb{A}\, \underline{U}_\star = \underline{F}_u. \\ \text{Pressure solver:} & \text{Solve in } \underline{P} \text{ such that } \mathbb{K}\, \underline{P} = -\nu\, \lambda \mathbb{B}\, \underline{U}_\star. \\ \text{Correction:} & \text{Solve in } \delta \underline{U} \text{ such that } \nu\, \mathbb{A}\, \delta \underline{U} = \mathbb{B}^T\, \underline{P}. \\ \text{Update:} & \underline{U} = \delta \underline{U} + \underline{U}_\star. \end{array} \tag{7.7.29}$$

One can check easily that the above computed solution $(\underline{U}, \underline{P})$ solves (7.7.28). The pressure solver with matrix $\mathbb{K}$ is based on the Uzawa algorithm, which is the conjugate gradient algorithm in the context of the Stokes problem. It can be preconditioned by the inverse of the mass matrix associated to the discrete pressure $\mathbb{M}$ (see e.g. [205, Lemma 5.9]). Thanks to the uniform discrete inf-sup condition, the number of iterations of the conjugate gradient algorithm is independent of the meshsize. The matrices $\mathbb{M}$ and $\mathbb{B}$ are kept with all the P_{disc}^{k-1} degrees of freedom for the discrete pressure. To take into account the zero mean value constraint, at each iteration of the preconditioned conjugate gradient algorithm, the discrete pressure is first computed in $L^2(\Omega)$, then orthogonally projected in $L^2_{zmv}(\Omega)$. We use the Cholesky factorization (computed once and for all) to solve linear systems with matrix $\mathbb{A}$.

Consider next the $\mathbf{P}^k \times P_{disc}^{k-1}$ conforming discretization of the variational formulation (7.7.17) or (7.7.24). Let $(\psi_i)_{i=1}^{N_u}$ be the Lagrange basis of $\mathbf{V}_{0,h}^k$ and $(\phi_i)_{i=1}^{N_p}$ be the basis of Q_h^{k-1}. We set: $\underline{U} := (\underline{U}_i)_{i=1}^{N_u}$ where $\mathbf{u}_h := \sum_{i=1}^{N_u} \underline{U}_i \psi_i$ and $\underline{P} := (\underline{P}_i)_{i=1}^{N_p}$ where $p_h := \sum_{i=1}^{N_p} \underline{P}_i \phi_i$. We set $\underline{F}_u := (\underline{F}_{u,i})_{i=1}^{N_u}$ where $\underline{F}_{u,i} = \langle \mathbf{f}, \psi_i \rangle_{\mathbf{H}_0^1(\Omega)}$ and $\underline{F}_p := (\underline{F}_{p,i})_{i=1}^{N_p}$ where $\underline{F}_{p,i} = (z_{\mathbf{f}}, \phi_i)_{L^2(\Omega)}$, cf. (7.7.17), or $\underline{F}_{p,i} = (p_h^{old}, \phi_i)_{L^2(\Omega)}$, where the discrete pressure p_h^{old} is an approximation of $z_{\mathbf{f}}$, cf. (7.7.24). The linear system to be solved is

$$\left\{ \begin{array}{l} \text{Find } (\underline{U}, \underline{P}) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p} \text{ such that:} \\ \nu\lambda \mathbb{A}\, \underline{U} - \lambda \mathbb{B}^T \underline{P} = \lambda \underline{F}_u \\ \lambda \mathbb{B}\, \underline{U} + \nu^{-1} \mathbb{M}\, \underline{P} = \nu^{-1} \underline{F}_p \end{array} \right. . \tag{7.7.30}$$

In that case, we set $\mathbb{A}_\lambda = \mathbb{A} + \lambda \mathbb{B}^T \mathbb{M}^{-1} \mathbb{B} \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_u}$, which is automatically a symmetric positive definite matrix. To solve the coupled velocity-pressure problem (7.7.30), one relies on the two steps below :

$$\begin{array}{ll} \text{Velocity solver:} & \text{Solve in } \underline{U} \text{ such that } \nu\, \mathbb{A}_\lambda\, \underline{U} = \underline{F}_u + \mathbb{B}^T \mathbb{M}^{-1} \underline{F}_p. \\ \text{Pressure solver:} & \text{Solve in } \underline{P} \text{ such that } \mathbb{M}\, \underline{P} = \underline{F}_p - \nu\lambda \mathbb{B}\, \underline{U}. \end{array} \tag{7.7.31}$$

One can check easily that the above computed solution $(\underline{U}, \underline{P})$ solves (7.7.30). The matrix $\mathbb{A}_\lambda$ can be assembled easily since $\mathbb{M}$ is a block-diagonal matrix. Using $\mathbf{P}^1 \times P^0$ discretization, the computation of $\underline{P}$ is straightforward. Again, matrices $\mathbb{M}$ and $\mathbb{B}$ are kept with all the P_{disc}^{k-1} degrees of freedom for the discrete pressure. We proceed as before to take into account the zero mean value constraint.

Remark 7.7.5. *We recall that solving the linear system (7.7.28) via the solver (7.7.29) is not so straightforward, even with $\nu = 1$. Solving the linear system (7.7.30) is much easier.*

Settings

We consider Problem (7.2.1) with homogeneous or non homogeneous boundary conditions in $\Omega = (0,1)^2$. Let $(\mathbf{u}, p)$ be the exact solution, and $(\mathbf{u}_h, p_h)$ be the numerical solution. We compare numerical methods, showing how the coercive $\mathbf{P}^1 \times P^0$ formulation can be used in a post-processing step (i.e. solving (7.7.24) with p_h^{old} known), improving then the approximation of the discrete velocity. In Figures 7.10 to 7.19, and in Tables 7.10 to 7.14 we give the following names to our numerical methods:

- Method with Crouzeix-Raviart (CR): computations are made with the nonconforming Crouzeix-Raviart $\mathbf{P}^1_{nc} \times P^0$ formulation [121, Example 4], which is not a pressure robust method. We call p_{nc} the resulting discrete pressure.
- Method with exact pressure (EP): computations are made with the coercive $\mathbf{P}^1 \times P^0$ formulation (7.7.17) and $\lambda = 1$, knowing the exact pressure $z_{\mathbf{f}}$ (i.e. we solve Problem (7.7.17)).
- Method with post-processing (Post): computations are made in two steps. In a first step, we approximate the pressure by the CR-method. Then, in a post processing step, we use this numerical pressure as the source term in the EP-method (i.e. we solve Problem (7.7.24) with $p_h^{old} = p_{nc}$) and $\lambda = 1$ when $\nu = 1$, $\lambda = 10$ else. Unless the stopping criterion (7.7.16) is achieved, we iterate eight times the second step, updating p_h^{old} at each new iteration. The algorithm corresponds to Algorithm (7.7.12).

Let N_T be the number of triangles. For the velocity, the number of unknowns is of order $N_u \approx 3\,N_T$ for the CR-method and $N_u \approx N_T$ for the EP-method. For the pressure, the number of unknowns is $N_p = N_T - 1$ for both methods. As a consequence, there are roughly twice as many unknowns for the CR-method than there are for the EP-method. We report in Table 7.8 the number of unknowns "# dof" for the numerical tests.

h	# dof CR	# dof EP	# dof CR	# dof EP
1.00×10^{-1}	1 048	566	2 184	1 114
5.00×10^{-2}	4 376	2 270	8 064	4 074
2.50×10^{-2}	17 368	8 846	30 464	15 314
1.25×10^{-2}	67 816	34 230	117 664	58 994
6.25×10^{-3}	272 624	136 954	464 448	232 866

Table 7.8: Number of unknowns: first two test cases (left), last test case (right).

We propose three numerical examples based on manufactured solutions. Let $\mathcal{I}_{h,cg}$ be the interpolation operator from $\mathcal{C}^0(\overline{\Omega})$ to $\mathbf{V}^1_{0,h}$. We use the discrete L^2-error for the velocity and for the pressure:

$$\varepsilon_0^\nu(\mathbf{u}_h) = \|\mathcal{I}_h\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)}/\|(\mathbf{u},p)\|_{\mathcal{X},\nu}, \quad \varepsilon_0^\nu(p_h) = \nu^{-1}\|\Pi_{h,dg}p - p_h\|_{L^2(\Omega)}/\|(\mathbf{u},p)\|_{\mathcal{X},\nu} \tag{7.7.32}$$

where $\mathcal{I}_h$ is either equal $\mathcal{I}_{h,nc}$ (cf. definition 7.4.1) or $\mathcal{I}_{h,cg}$. For the EP-method and the post-method, we moreover compare the indicators ε_D^ν and ε_1^ν defined by:

$$\varepsilon_1^\nu = \|\,\boldsymbol{\nabla}(\mathcal{I}_{h,cg}\mathbf{u} - \mathbf{u}_h)\|_{\mathbb{L}^2(\Omega)}/\|(\mathbf{u},p)\|_{\mathcal{X},\nu}, \quad \varepsilon_D^\nu = \|\,\mathrm{div}\,\mathbf{u}_h\|_{L^2(\Omega)}/\|(\mathbf{u},p)\|_{\mathcal{X},\nu}. \tag{7.7.33}$$

We first plot the results as error curves for both the velocity and the pressure as a function of the meshsize, and we give the numerically observed average convergence rates. Second, we report the elapsed CPU times for the CR-method and the Post-method. In the tables, column "overhead" indicates the ratio

between the cost of the second (post processing) step for the Post-method and the cost of the first step (CR-method). Finally, we plot the errors of the two methods as a function of the elapsed CPU times of the resolution algorithm.

For the Post-method "Post" plots represent the errors after a single iteration of the second step and "Post-8" plots represent the errors after iterating the second step eight times, updating p_h^{old} at each new iteration, cf. Algorithm (7.7.12).

Remark 7.7.6. *The use of discrete mixed formulations with* $\mathbf{H}(\mathrm{div})$*-conforming projection of the test function on the right-hand side leads to a pressure robust discrete velocity [194, 66]. It has been proven to be an efficient numerical solution to solve Stokes problem when the parameter* ν *is small, however the* $\mathbf{H}(\mathrm{div})$*-conforming discrete spaces must be chosen with care. The Post-method can also be used in that case to obtain an* $\mathbf{H}^1$*-conforming reconstruction of the discrete velocity, see §7.7.5.*

Regular manufactured solutions

We postulate that for the EP-method and the Post-method, $\varepsilon_0^\nu(\mathbf{u}_h) \leq C_0\, h^2$, where C_0 is independent of the meshsize. The error estimates for the CR-method are given by [121, Theorem 3, Theorem 4, Theorem 6]. Notice that the norm $\|(\cdot,\cdot)\|_{\mathcal{X},\nu}$ depends on the viscosity ν (7.2.3) and that the error estimates on the pressure and the velocity are linked. We check the convergence rates and the viscosity dependency of the error estimates. The first test enters the framework of Proposition 7.7.4.

- **Test case with a linear velocity: Figures 7.9-7.12, Tables 7.9-7.10.**

We consider Problem $(7.2.1)_{NH}$ with $\mathbf{f} = \nabla\, p$, where: $\mathbf{u} = (-y, x)^T$ and $p = x^3 + y^3 - 1/2$.

The number of unknowns are reported in Table 7.8 (left).

As expected in (7.7.19), the EP-method returns $\varepsilon_0^\nu(\mathbf{u}_h) = \mathcal{O}(10^{-15})$, $\varepsilon_0^\nu(p_h) = \mathcal{O}(10^{-14})$ for $\nu = 1$; and $\varepsilon_0^\nu(\mathbf{u}_h) = \mathcal{O}(10^{-18})$, $\varepsilon_0^\nu(p_h) = \mathcal{O}(10^{-17})$ for $\nu = 10^{-6}$, so the errors are not reported.

Figure 7.9 (resp. 7.10) shows the discrete error values $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right) plotted against the meshsize for $\nu = 1$ (resp. $\nu = 10^{-6}$). Notice that the Post-method improves the approximation initially computed with the CR-method.

Figure 7.11 shows the discrete error values ε_1^ν and ε_D^ν plotted against the meshsize for $\nu = 10^{-6}$. Notice that the ratio $\varepsilon_D^\nu/\varepsilon_1^\nu$ is of order 0.5 for the Post-method with a single iteration and of order 0.1 for the Post-method with eight iterations.

Table 7.9 shows the average convergence rates for the velocity, $\tau_{\mathbf{u}}$ and the pressure, τ_p. For the CR-method, the average convergence rate for the pressure, τ_p is better than expected, possibly because the source term is a polynomial of degree 3 which gradient is numerically exactly integrated.

Table 7.10 shows the CPU times for the CR-method and the Post-method with either one, or eight, iterations. With our stopping criterion, around 30 iterations of the preconditioned conjugate gradient algorithm are needed to solve the pressure solver of algorithm (7.7.29) with the CR-method: this is consistent with Remark 7.7.5. By design, the Post-method which includes the CR-method as its first step requires more CPU time. However, we notice that the CPU time of the second (post processing) step, the overhead, is only a small fraction of the first step, and also that it decreases dramatically as the meshsize decreases. This can be explained by the fact that on the one hand, there are fewer unknowns and, on the other hand, Algorithm (7.7.31) is faster than Algorithm (7.7.29) (cf. Remark 7.7.5). For the Post-method, it seems worth doing eight iterations, especially when the meshsize is small.

Figure 7.12 shows the discrete error values $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right) plotted against the CPU time for $\nu = 10^{-6}$. To reach $\varepsilon_0^\nu(\mathbf{u}_h) \lesssim 2 \times 10^{-5}$, the required CPU time is $0.5\, s$ for the Post-method with one iteration and less than $0.01\, s$ for the Post-method with eight iterations, to be compared with $20\, s$ for the CR-method.

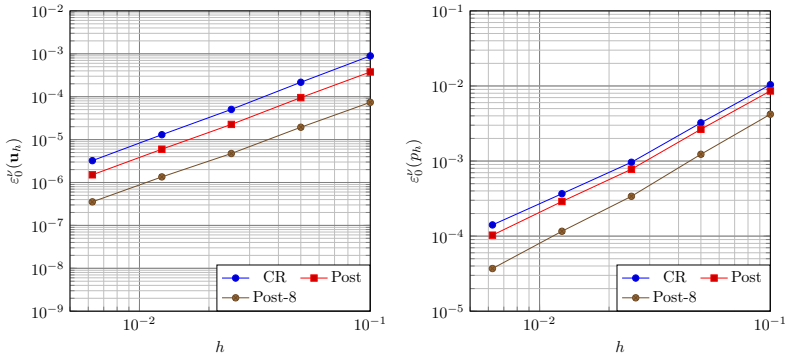


Figure 7.9: Linear velocity, $\nu = 1$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

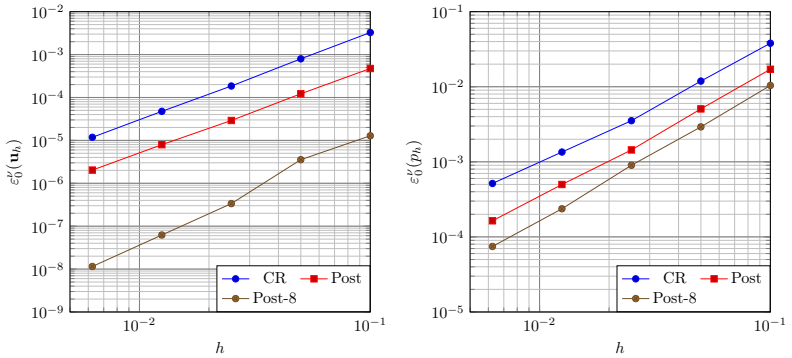


Figure 7.10: Linear velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

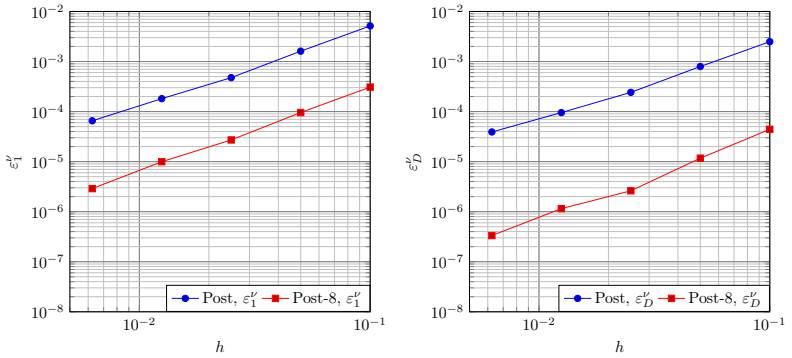


Figure 7.11: Linear velocity, $\nu = 10^{-6}$. Plots of ε_1^ν (left) and ε_D^ν (right).

- **Test case with a sinusoidal solution: Figures 7.13-7.16, Tables 7.12-7.11**

We consider Problem $(7.2.1)_H$ with $\mathbf{f} = -\nu\,\boldsymbol{\Delta}\,\mathbf{u} + \nabla\, p$, where:

$$\mathbf{u} = \begin{pmatrix} (1 - \cos(2\,\pi\,x))\,\sin(2\,\pi\,y) \\ (\cos(2\,\pi\,y) - 1)\,\sin(2\,\pi\,x) \end{pmatrix} \text{ and } p = \sin(2\,\pi\,x)\,\sin(2\,\pi\,y).$$

The number of unknowns are reported in Table 7.8 (left).

Figure 7.13 shows $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right) plotted against the meshsize, for $\nu = 1$. In that case, for the Post-method, there is no need doing eight iterations, so we give numerical results for computations with a single iteration only. The Post-method and the EP-method give similar velocity errors. The CR-method and the Post-method give similar pressure errors. For a given meshsize h, the

ν	τ	CR	Post	Post-8
1	$\tau_{\mathbf{u}}$	2.03	1.99	1.93
	τ_p	1.55	1.59	1.71
10^{-6}	$\tau_{\mathbf{u}}$	2.03	1.97	2.53
	τ_p	1.55	1.68	1.78

Table 7.9: Linear velocity. Average convergence rates.

h	CPU CR	CPU Post	overhead	CPU Post-8	overhead
1.00×10^{-1}	4.34×10^{-3}	5.08×10^{-3}	17 %	7.40×10^{-3}	70 %
5.00×10^{-2}	2.32×10^{-2}	2.61×10^{-2}	13 %	4.46×10^{-2}	92 %
2.50×10^{-2}	2.51×10^{-1}	2.69×10^{-1}	7.2 %	3.65×10^{-1}	45 %
1.25×10^{-2}	$3.00 \times 10^{+0}$	$3.07 \times 10^{+0}$	2.3 %	$3.54 \times 10^{+0}$	10 %
6.25×10^{-3}	$4.89 \times 10^{+1}$	$4.94 \times 10^{+1}$	1.0 %	$5.20 \times 10^{+1}$	6.3 %

Table 7.10: Linear velocity, $\nu = 10^{-6}$. CPU time (s).

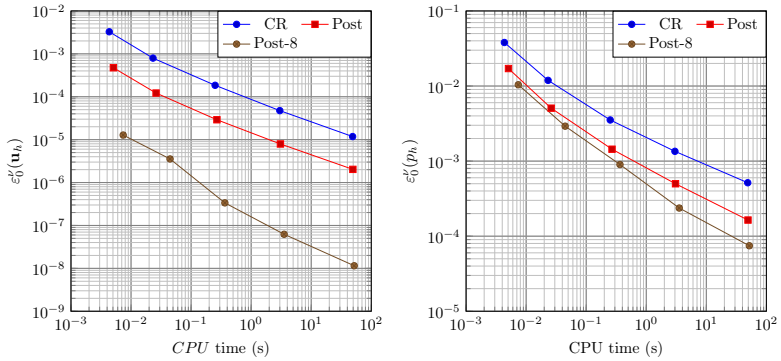


Figure 7.12: Linear velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right) against CPU time (s).

velocity error is smaller for the EP-method and the Post-method than for the CR-method; and the pressure error is smaller for the CR-method than for the EP-method, but it is obtained at a higher cost. Notice that the Post-method reduces the velocity error without worsening the pressure error.

Figure 7.14 shows $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right) plotted against the meshsize, now for $\nu = 10^{-6}$. Both post-processings improve the initial computation. Even with eight iterations, the overhead cost remains affordable, especially when the meshsize is small.
The EP-method gives much smaller velocity and pressure errors than the CR-method since it is a pressure robust method. For a given meshsize h, we note that the Post-method allows to reduce the velocity error by a factor larger than 10. The pressure error of the Post-method using one iteration is close to that of the CR-method, while it is greatly improved using eight iterations.

Figure 7.15 shows the discrete error values ε_1^ν and ε_D^ν plotted against the meshsize for $\nu = 10^{-6}$. Iterating several times yields again a significant decrease of the indicators.

Table 7.11 shows the average convergence rates for the velocity, $\tau_{\mathbf{u}}$ and the pressure, τ_p. In the case $\nu = 1$, the average convergence rates are as expected. In the case $\nu = 10^{-6}$, the average convergence rate for the velocity, $\tau_{\mathbf{u}}$, is better than expected for the EP-method and the Post-method, and the average convergence rate for the pressure, τ_p, is better than expected for the Post-method, probably because the asymptotic convergence regime has not been reached.

Table 7.12 below shows the CPU times for the CR-method and the Post-method with either one, or eight, iterations. As we have used the same meshes, the computation times are similar to those in Table 7.10, showing once again that the overhead cost decreases dramatically as the mesh size decreases.

Figure 7.16 shows $\varepsilon_0^\nu(\mathbf{u}_h)$ (right) and $\varepsilon_0^\nu(p_h)$ (left) plotted against the CPU time for the CR-method and the Post-method with a single iteration ("Post" plot) or eight iterations ("Post-8" plot). In the same way as for the first test, we note that to reach $\varepsilon_0^\nu(\mathbf{u}_h) \lesssim 5 \times 10^{-6}$, the required CPU time is $0.2\,s$ for the Post-method with one iteration and less than $0.02\,s$ for the Post-method with eight iterations, to be compared with the CR-method which does not reach this threshold.

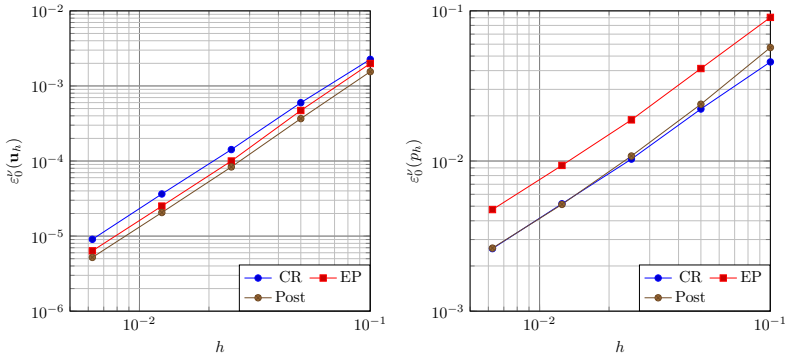


Figure 7.13: Sinusoidal velocity, $\nu = 1$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

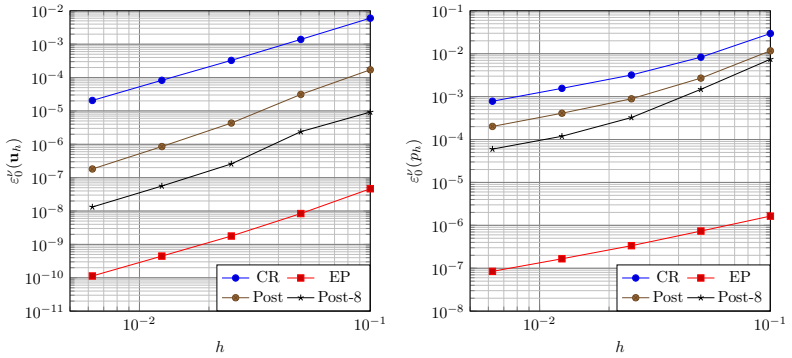


Figure 7.14: Sinusoidal velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

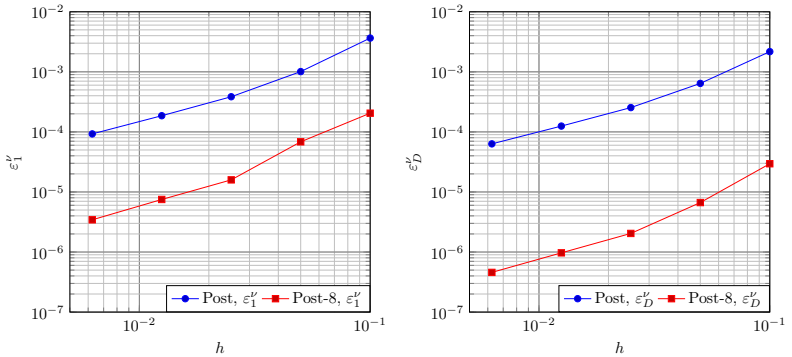


Figure 7.15: Sinusoidal velocity, $\nu = 10^{-6}$. Plots of ε_1^ν (left) and ε_D^ν (right).

ν	τ	CR	EP	Post	Post-8
1	$\tau_{\mathbf{u}}$	1.99	2.07	2.06	–
	τ_p	1.03	1.06	1.11	–
10^{-6}	$\tau_{\mathbf{u}}$	2.05	2.17	2.47	2.36
	τ_p	1.31	1.07	1.47	1.74

Table 7.11: Sinusoidal velocity. Average convergence rates.

h	CPU CR	CPU Post	overhead	CPU Post-8	overhead
1.00×10^{-1}	3.85×10^{-3}	4.64×10^{-3}	21 %	7.06×10^{-3}	83 %
5.00×10^{-2}	2.22×10^{-2}	2.51×10^{-2}	13 %	4.27×10^{-2}	92 %
2.50×10^{-2}	2.43×10^{-1}	2.58×10^{-1}	6.2 %	3.45×10^{-1}	42 %
1.25×10^{-2}	$4.27 \times 10^{+0}$	$4.34 \times 10^{+0}$	1.6 %	$4.78 \times 10^{+0}$	12 %
6.25×10^{-3}	$4.53 \times 10^{+1}$	$4.57 \times 10^{+1}$	0.7 %	$4.82 \times 10^{+1}$	6.4 %

Table 7.12: Sinusoidal velocity, $\nu = 10^{-6}$. CPU time (s).

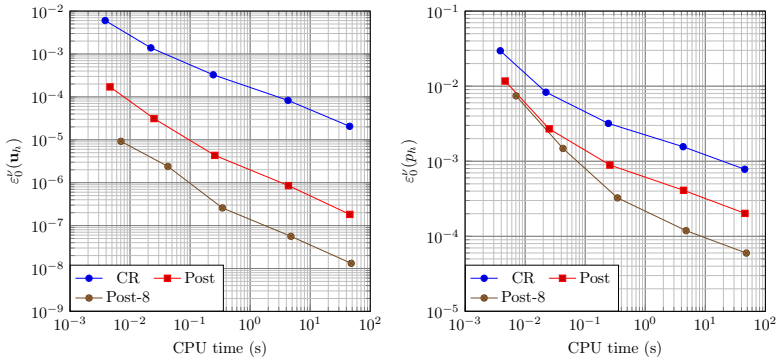


Figure 7.16: Sinusoidal velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ and $\varepsilon_0^\nu(p_h)$ against CPU time (s).

- **Comment on the algorithm**

At this stage, we noticed that eight iterations was a good compromise. In order to design an optimized algorithm, another stopping criterion could be set by comparing two successive computations of p_h^{old}, i.e. comparing $|\delta \underline{P}^n|_{\mathbb{M}}$ to $|\delta \underline{P}^{n-1}|_{\mathbb{M}}$. The choice of λ could also be further optimized, depending on ν and on the meshsize.

- **Some observations**

The coercive $\mathbf{P}^1 \times P^0$ formulation (EP-method) gives *pressure robust* results, the obvious limitation being that it requires to know explicitly the potential of the gradient part of the source term. If it is not known, the two step method greatly reduces the velocity error, compared with the calculation carried out using the Crouzeix-Raviart $\mathbf{P}^1_{nc} \times P^0$ formulation (CR-method). Moreover, using the second (post processing) step *iteratively* greatly improves the initial result. Finally, the reduction factor is greater the smaller ν is.

Low regularity manufactured solution: Figures 7.17-7.19, Tables 7.13-7.14

Last, we consider Problem $(7.2.1)_{NH}$ with a low regularity solution. Let (ρ, θ) be the polar coordinates centered in $(0.5, 0.5)$. Let $\alpha = 0.45$. We set $\mathbf{f} = -\nu \boldsymbol{\Delta} \mathbf{u} + \nabla p$ where $(\mathbf{u}, p) = (\rho^\alpha \mathbf{e}_\theta, \rho - \int_\Omega \rho)$. Results are given for $\nu = 10^{-6}$. The number of unknowns are reported in Table 7.8 (right). We used a refined mesh around $(0.5, 0.5)$, where the solution is of low regularity.

Figure 7.17 shows the discrete error values $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right) plotted against the meshsize. We remark that EP-method shows far better results than the CR-method, and that the Post-method allows again to improve the approximation of the CR-method.

Figure 7.18 shows the discrete error values ε_1^ν (left) and ε_D^ν (right) plotted against the meshsize. We have $\varepsilon_D^\nu / \varepsilon_1^\nu \approx 0.5$ for the Post-method with a single iteration while $\varepsilon_D^\nu / \varepsilon_1^\nu \approx 0.1$ or the Post-method with eight iterations.

Table 7.13 shows the averaged convergence rates between the successive meshes. For the EP-method and the CR-method, we postulate that, asymptotically, $\tau_{\mathbf{u}} = 1 + \alpha$ and $\tau_p = \alpha$. In both cases, the obtained convergence rates are better than expected, which suggest that the asymptotic convergence regime is not reached.

Table 7.14 below shows the CPU times for the CR-method and the Post-method with eight iterations. Again, the overhead cost decreases sharply with the meshsize.

Figure 7.19 shows the discrete error values $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right) against the CPU time for the CR-method and the Post-method with eight iterations. To reach $\varepsilon_0^\nu(\mathbf{u}_h) \lesssim 10^{-5}$, the required CPU time is $0.02\,s$ for the Post-method with eight iterations, to be compared with more than $100\,s$ for the CR-method.

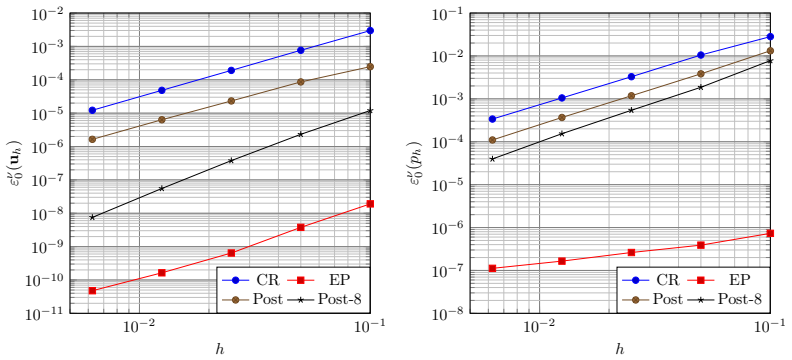


Figure 7.17: Low regularity velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

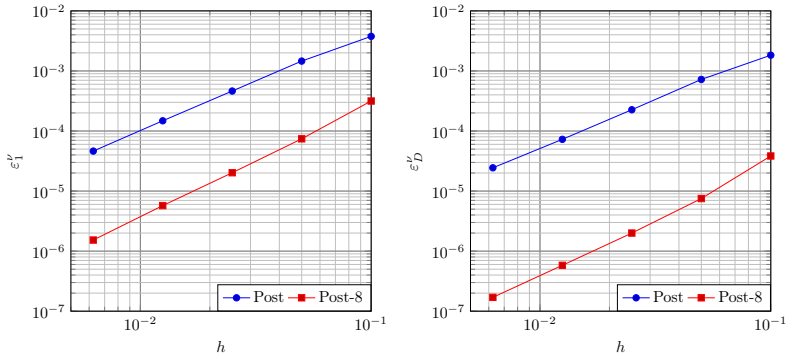


Figure 7.18: Low regularity velocity, $\nu = 10^{-6}$. Plots of ε_1^ν (left) and ε_D^ν (right).

τ	CR	EP	Post-8
$\tau_{\mathbf{u}}$	2.0	2.2	2.8
τ_p	1.6	0.7	1.9

Table 7.13: low regularity velocity, $\nu = 10^{-6}$. Averaged convergence rates.

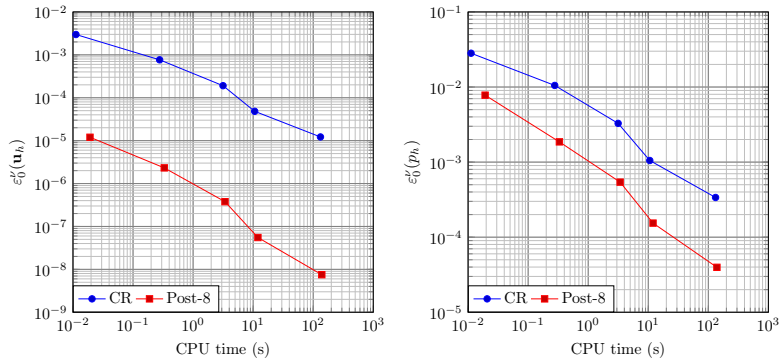


Figure 7.19: Low regularity velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ and $\varepsilon_0^\nu(p_h)$ against CPU time (s).

Numerical results using Raviart-Thomas projection

The use of Crouzeix-Raviart $\mathbf{P}_{nc}^1 \times P^0$ formulation with $\mathbf{H}(\mathrm{div})$-conforming projection of the test function on the right-hand side leads to a pressure robust discrete velocity [194, 66]. In this case too, we can use the Post-method to obtain a precise $\mathbf{H}^1$-conforming approximation of the discrete velocity. In Figures 7.20-7.22, we represent $\varepsilon_0^\nu(\mathbf{u}_h)$ and $\varepsilon_0^\nu(p_h)$ against the meshsize for our three tests. For the CR-method, we use the lowest order Raviart-Thomas projection [194, §3.2] to compute p_h^{old}, which is then used in the Post-method with $\lambda = 1$. For the first test (linear velocity), we obtain $\varepsilon_0^\nu(\mathbf{u}_h) \lesssim 10^{-13}$ and $\varepsilon_0^\nu(p_h) \lesssim 10^{-12}$, and regardless, we observe that the errors (close to machine precision for the CR-method) do not deteriorate when the Post-method is applied. The pressure error $\varepsilon_0^\nu(p_h)$ remains unchanged for the other two tests. The approximation of the velocity is improved for all three tests.

h	CPU CR	CPU Post-8	overhead
1.00×10^{-1}	1.12×10^{-2}	1.92×10^{-2}	71 %
5.00×10^{-2}	2.77×10^{-1}	3.31×10^{-1}	19 %
2.50×10^{-2}	$3.15 \times 10^{+0}$	$3.40 \times 10^{+0}$	7.9 %
1.25×10^{-2}	$1.06 \times 10^{+1}$	$1.20 \times 10^{+1}$	13 %
6.25×10^{-3}	$1.32 \times 10^{+2}$	$1.39 \times 10^{+2}$	5 %

Table 7.14: Low regularity velocity, $\nu = 10^{-6}$. CPU time (s).

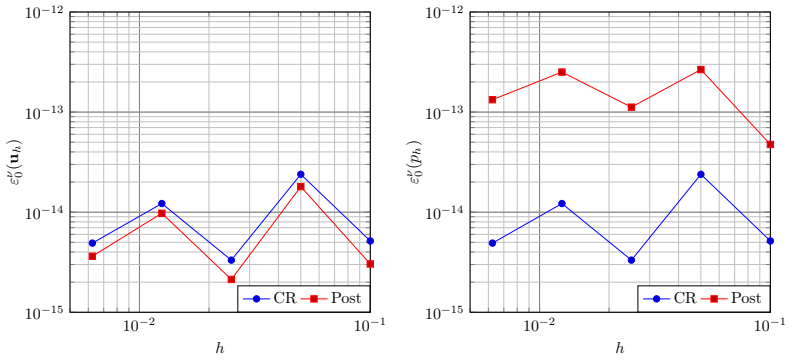


Figure 7.20: Linear velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

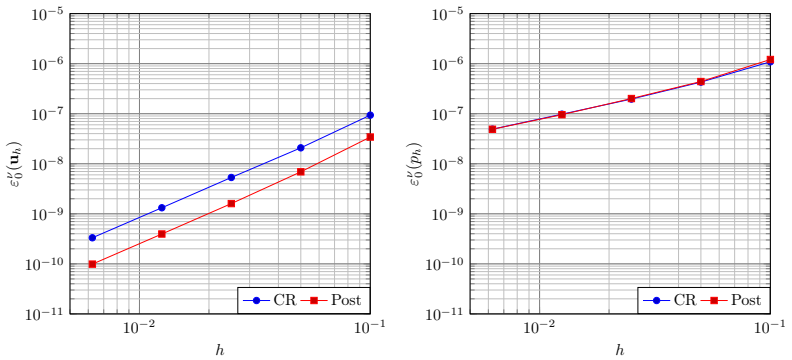


Figure 7.21: Sinusoidal velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

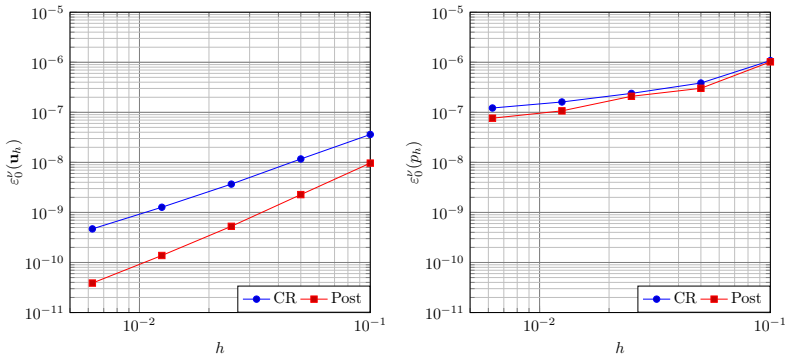


Figure 7.22: Low regularity velocity, $\nu = 10^{-6}$. Plots of $\varepsilon_0^\nu(\mathbf{u}_h)$ (left) and $\varepsilon_0^\nu(p_h)$ (right).

Chapter 8

Perspectives

In Chapter 5, we studied three different electromagnetic models. Concerning the magnetostatic system, with the help of **explicit T-coercivity**, we recovered a well-known perturbed variational formulation. We carried analysis the numerical analysis of its discretization with edge finite elements with more general assumptions. We also studied the mixed variational formulations of our three models with the **weighted regularization method**. This method allows to use continuous vector finite elements, which can be of interest in the case of Vlasov-Maxwell coupling. We now wish to prove the discrete inf-sup condition.

In Chapter 6, we carried out numerical analysis of the discretization of the neutron diffusion equation with the DD+L^2-jumps method. As I changed of department, this work was taken over by François Madiot and Patrick Ciarlet. The PhD thesis of Léandre Giret was followed by the post-doc of Minh-Hieu Do, on ***a posteriori* estimators** for this equation [132, 96]. Mario Gervais is currently working on his PhD thesis on extending these *a posteriori* estimators to the DD+L^2-jumps method, and to the multigroup SP_N equations [160]. These estimators are plugged into the APOLLO3® code, allowing to improve its performance.

Using **explicit T-coercivity** for the neutron diffusion equation allows to exhibit an **efficient preconditionner**. We plan to extend the study first to the DD+L^2-jumps method and then, to other models.

We also studied a simplified thermal-hydraulics and neutronics coupling, based on the pressurized water reactor model. We now turn our attention to the molten-salt reactor model, for which we need to take into account the Bateman equations, describing the evolution of the isotopic concentrations.

In Chapter 7, we detailed the spatial discretization of the TrioCFD code. First, in Section 7.4, we studied nonconforming finite element. In Section 7.5, we proved the stability and computed errors estimates $\mathbf{P}^1_{nc} - (P^0 + P^1)$ discretization. In Section 7.6, we proposed an alternative scheme for the discretization of the pressure. In Section 7.7, we proposed and analysed **a new variational formulation of the Stokes problem using explicit T-coercivity**. This variational formulation is coercive, and can be discretized with the $\mathbf{P}^k \times P^{k-1}_{disc}$ finite element for all $k \geq 1$. To solve the linear system resulting from the discretization, we need to know the pressure, or at least some approximation of it, which in our numerical tests is the discrete pressure obtained using the classical nonconforming method with the Crouzeix-Raviart $\mathbf{P}^1_{nc} \times P^0$ finite element. This two-step method improves the numerical results by notably reducing the errors obtained after the use of the classical method, especially when the viscosity is small.

In these three main Chapters, for the mixed problems, we used T-coercivity framework to find an explicit realization of the inf-sup condition for the all-by-one bilinear forms. It provides useful insight

on the design of suitable discretizations. Furthermore, the T operator may be used to express a coercive variational formulation. Hence, although it does not provide any new theoretical results, T-coercivity framework helps for the study and the approximation of mixed variational formulations [26, 95].

Concerning TrioCFD code, our aim is to enhance its capabilities. I am involved in supervising two PhD students, working on aspects of this topic.

In order to handle triangulation with general meshes and to improve the simulation of vortices, we want to use **discontinuous Galerkin finite elements**. Currently working on her PhD thesis, **Mayssa Mroueh** is exploring different numerical schemes, based on the Symmetric Interior Penalty method to solve the Stokes problem. Classically, the velocity is discretized with $\mathbf{P}^k_{disc}$ finite elements and the pressure with P^{k-1}_{disc} finite elements [162], where $k > 1$. One can also discretize the pressure with P^k_{disc} finite elements, with the help of a stabilization term [130, Section 6.1]. Mayssa proved that the stability is achieved also when the pressure is discretized with P^{k+1}_{disc} finite elements. Numerical results showed that this new scheme may be interesting when the viscosity is small [P9, Section 6.2]. On the other hand, we computed ***a priori* error estimates** for those different schemes when the source term is of **low-regularity** [W2]. Mayssa is also studying DG schemes using **exponential basis functions** instead of polynomial ones. Indeed, these functions provide a more accurate model of convective flows. Implementation of the DG schemes in the TrioCFD code is currently underway. We intend now to study *a posteriori error* estimators, in order to propose latter adaptive *hp* strategies. Another direction it to extend the T-coercivity theory and study to **Oseen problem** (i.e. Stokes problem with an advection term).

The TrioCFD code can simulate single-phase flows, and it contains a front-tracking algorithm to simulate two-phase flows by interface tracking, using two meshes: a moving mesh for the interface, and a fixed mesh on which fluid velocity, temperature and pressure are calculated. The Navier-Stokes equations are solved in each phase, and the coupling between phases is modelled by a surface force field. This module is being used more and more, but some calculations are too costly. In particular, when we want to simulate the interface of a vortex, we have to refine drastically the mesh in the neighbourhood of the vortex extremity. During the work-study period, **Paulin Dempowo-Fouelefack** analysed and helped improve the front-tracking algorithm. He will now start a PhD thesis, during which we plan to adapt the singular complement method proposed by Lefebvre-Lepot and Nabet [190] to simulate a flow containing rigid particles. **This method consists in taking into account explicitly the singular part of the fluid velocity between particles, which would need a very fine mesh to be approximated accurately.** Only the regular part and a coefficient are computed on a rough mesh. In their context, the authors show that for a given precision, this method allowed to use a coarser mesh. We would like to adapt this strategy to take into account the interface singularities.

Appendix A

Neutronics

A.1 Inverse power algorithm

Consider Problem (6.1.14). We look for the eigenpair $((\mathbf{p},\phi),k_{eff})$ where k_{eff} is the inverse of the smallest eigenvalue, so that we can use the so-called inverse power algorithm. After some initial guess is provided, at iteration number $m+1$, we deduce $(\mathbf{p}^{m+1},\phi^{m+1},k_{eff}^{m+1})$ from $(\mathbf{p}^{m},\phi^{m},k_{eff}^{m})$ by solving Equations (6.1.14) with a source term. Set in a domain $\mathcal{R}$, the inverse power iteration algorithm[1] reads (we choose below zero flux boundary conditions on $\partial\mathcal{R}$):

Set $(\mathbf{p}^0,\phi^0,k_{eff}^0)$, $m=0$.

Until convergence, do: $m \leftarrow m+1$

Solve:

$$\left\{\begin{array}{rcll} D^{-1}\,\mathbf{p}^{m+1}+\nabla\,\phi^{m+1} &=& 0 & \text{in } \mathcal{R},\\ \text{div }\mathbf{p}^{m+1}+\Sigma_a\,\phi^{m+1} &=& \dfrac{1}{k_{eff}^m}\underline{\nu\Sigma}_f\,\phi^m & \text{in } \mathcal{R},\\ \phi^{m+1} &=& 0 & \text{on } \partial\mathcal{R}. \end{array}\right. \tag{A.1.1}$$

Compute: $k_{eff}^{m+1}=k_{eff}^{m}\,\dfrac{\displaystyle\int_{\mathcal{R}}(\,\underline{\nu\Sigma}\,\phi^{m+1}\,)^2}{\displaystyle\int_{\mathcal{R}}(\,\underline{\nu\Sigma}\,\phi^{m+1}\,\underline{\nu\Sigma}\,\phi^{m}\,)}$.

End

Above, the Eqs. (A.1.1) with unknowns $(\mathbf{p}^{m+1},\phi^{m+1})$ model the so-called *source solver* (here, the right-hand side is a function of ϕ^m). The updated value k_{eff}^{m+1} is inferred as follows: assuming that $\text{div }\mathbf{p}^{m+1}+\Sigma_a\,\phi^{m+1}=(k_{eff}^{m+1})^{-1}\underline{\nu\Sigma}\,\phi^{m+1}$, one can write $(k_{eff}^{m+1})^{-1}\underline{\nu\Sigma}\,\phi^{m+1}=(k_{eff}^{m})^{-1}\underline{\nu\Sigma}\,\phi^{m}$ and, multiplying this equation by $\underline{\nu\Sigma}\,\phi^{m+1}$ and integrating over the domain of computation $\mathcal{R}$, we obtain the last equation of (A.1.1). The convergence criterion is usually set on $|k_{eff}^{m+1}-k_{eff}^{m}|$ and $\int_{\mathcal{R}}(\,\underline{\nu\Sigma}\,\phi^{m+1}-\underline{\nu\Sigma}\,\phi^{m}\,)^2$. The inverse power iterations are called the outer iterations in opposition to the inner iterations, which correspond to the iterations of the source solver. As a matter of fact, one can choose an iterative source solver.

[1] This algorithm is described as well in [22] (algorithm 1, p. 2007).

A.2 Proof of Theorem 6.5.1

We note $\forall (i,j) \in \mathcal{IJ}$, $\alpha_{ij} := \alpha_{ij}^i = \alpha_{ij}^j > 0$. We introduce the sequence of errors $\Big((\mathbf{e}_i^\ell, e_i^\ell)_{\ell \in \mathbb{N}} \Big)_{i \in \mathcal{I}_\mathcal{R}}$, where: $\mathbf{e}_i^\ell := \mathbf{p}_i^\ell - \mathbf{p}_i \in \mathbf{H}(\mathrm{div}, \mathcal{R}_i)$ and $e_i^\ell := \phi_i^\ell - \phi_i \in V_i$. The sequence $\Big((\mathbf{e}_i^\ell, e_i^\ell)_{\ell \in \mathbb{N}} \Big)_{i \in \mathcal{I}_\mathcal{R}}$ satisfy for all $i \in \mathcal{I}_{\mathcal{R}_i}$, and for all $j \in \mathcal{I}_{\mathcal{R}_i}$:

$$\left\{ \begin{array}{lrcll} (0) & \mathrm{div}\, \mathbf{e}_i^\ell + \Sigma_{a,i}\, e_i^\ell & = & 0 & \text{in } \mathcal{R}_i, \\ (i) & (D_i)^{-1}\, \mathbf{e}_i^\ell + \nabla\, e_i^\ell & = & 0 & \text{in } \mathcal{R}_i, \\ (ii) & -\mathbf{e}_i^\ell \cdot \mathbf{n}_i + \alpha_{ij} e_i^\ell & = & \mathbf{e}_j^{\ell-1} \cdot \mathbf{n}_j + \alpha_{ij} e_j^{\ell-1} & \text{on } \Gamma_{ij}. \end{array} \right. \tag{A.2.1}$$

By induction on ℓ, we check that $\mathbf{e}_i^\ell \cdot \mathbf{n}_{i|\Gamma_{ij}} \in L^2(\Gamma_{ij})$. Multiplying (A.2.1)-(0) by e_i^ℓ and integrating on $\mathcal{R}_i$ gives:

$$\int_{\mathcal{R}_i} \mathrm{div}\, \mathbf{e}_i^\ell\, e_i^\ell + \|(\Sigma_{a,i})^{\frac{1}{2}}\, e_i^\ell\|^2_{L^2(\mathcal{R}_i)} = 0. \tag{A.2.2}$$

Multiplying (A.2.1)-(i) by $\mathbf{e}_i^\ell$, integrating on $\mathcal{R}_i$, using Eq. (4.2.1), we have:

$$\|(D_i)^{-\frac{1}{2}}\, \mathbf{e}_i^\ell\|^2_{L^2(\mathcal{R}_i)} - \int_{\mathcal{R}_i} \mathrm{div}\, \mathbf{e}_i^\ell\, e_i^\ell = - \sum_{j \in \mathcal{I}_{\mathcal{R}_i}} \int_{\Gamma_{ij}} \mathbf{e}_i^\ell \cdot \mathbf{n}_i\, e_i^\ell. \tag{A.2.3}$$

To bound the integrals on Γ_{ij}, we use that $AB = ((\alpha A + B)^2 - (\alpha A - B)^2)/(4\alpha)$ with $\alpha = \alpha_{ij}$; $A = e_i^\ell$ and $B = \mathbf{e}_i^\ell \cdot \mathbf{n}_i$ a.e. on Γ_{ij}; and the transmission conditions on Γ_{ij} (A.2.1)-(ii). Then we have for all $i \in \mathcal{I}_\mathcal{R}$ and for all $j \in \mathcal{I}_{\mathcal{R}_i}$:

$$\int_{\Gamma_{ij}} \mathbf{e}_i^\ell \cdot \mathbf{n}_i\, e_i^\ell = \Big(\|\mathbf{e}_i^\ell \cdot \mathbf{n}_i + \alpha_{ij} e_i^\ell\|^2_{L^2(\Gamma_{ij})} - \|\mathbf{e}_j^{\ell-1} \cdot \mathbf{n}_j + \alpha_{ij} e_j^{\ell-1}\|^2_{L^2(\Gamma_{ij})} \Big) / (4\, \alpha_{ij}). \tag{A.2.4}$$

Summing (A.2.2) and (A.2.3), and using (A.2.4) we then obtain for $i \in \mathcal{I}_\mathcal{R}$:

$$\|(\Sigma_{a,i})^{\frac{1}{2}}\, e_i^\ell\|^2_{L^2(\mathcal{R}_i)} + \|(D_i)^{-\frac{1}{2}}\, \mathbf{e}_i^\ell\|^2_{L^2(\mathcal{R}_i)} = \sum_{j \in \mathcal{I}_{\mathcal{R}_i}} \Big(\|\mathbf{e}_j^{\ell-1} \cdot \mathbf{n}_j + \alpha_{ij} e_j^{\ell-1}\|^2_{L^2(\Gamma_{ij})} - \|\mathbf{e}_i^\ell \cdot \mathbf{n}_i + \alpha_{ij} e_i^\ell\|^2_{L^2(\Gamma_{ij})} \Big) / (4\, \alpha_{ij}).$$

Summing the contributions of each sub-domain, we obtain for $\ell \in \mathbb{N}^*$:

$$\begin{aligned} & \sum_{i \in \mathcal{I}_\mathcal{R}} \Big(\|(\Sigma_{a,i})^{\frac{1}{2}}\, e_i^\ell\|^2_{L^2(\mathcal{R}_i)} + \|(D_i)^{-\frac{1}{2}}\, \mathbf{e}_i^\ell\|^2_{L^2(\mathcal{R}_i)} \Big) \\ = & \sum_{i \in \mathcal{I}_\mathcal{R}} \sum_{j \in \mathcal{I}_{\mathcal{R}_i}} \Big(\|\mathbf{e}_j^{\ell-1} \cdot \mathbf{n}_j + \alpha_{ij} e_j^{\ell-1}\|^2_{L^2(\Gamma_{ij})} - \|\mathbf{e}_i^\ell \cdot \mathbf{n}_i + \alpha_{ij} e_i^\ell\|^2_{L^2(\Gamma_{ij})} \Big) / (4\, \alpha_{ij}), \\ = & \sum_{(i,j) \in \mathcal{IJ}} \Big(\|\mathbf{e}_j^{\ell-1} \cdot \mathbf{n}_j + \alpha_{ij} e_j^{\ell-1}\|^2_{L^2(\Gamma_{ij})} - \|\mathbf{e}_j^\ell \cdot \mathbf{n}_j + \alpha_{ij} e_j^\ell\|^2_{L^2(\Gamma_{ij})} \cdot \\ & \qquad \|\mathbf{e}_i^{\ell-1} \cdot \mathbf{n}_i + \alpha_{ij} e_i^{\ell-1}\|^2_{L^2(\Gamma_{ij})} - \|\mathbf{e}_i^\ell \cdot \mathbf{n}_i + \alpha_{ij} e_i^\ell\|^2_{L^2(\Gamma_{ij})} \Big) / (4\, \alpha_{ij}) \end{aligned}$$

By summing over the iterations $\ell \in \{1, \cdots, N\}$, the contributions to successive iterations $\ell - 1$, ℓ cancel each other out for $\ell \in \{1, \cdots, N-1\}$ and it holds:

$$\begin{aligned} \sum_{\ell=1}^{N} \sum_{i \in \mathcal{I}_\mathcal{R}} \Big(\|(\Sigma_{a,i})^{\frac{1}{2}}\, e_i^\ell\|^2_{L^2(\mathcal{R}_i)} + \|(D_i)^{-\frac{1}{2}}\, \mathbf{e}_i^\ell\|^2_{L^2(\mathcal{R}_i)} \Big) + \sum_{(i,j) \in \mathcal{IJ}} \|\mathbf{e}_i^N \cdot \mathbf{n}_i + \alpha_{ij} e_i^N\|^2_{L^2(\Gamma_{ij})} / (4\, \alpha_{ij}) \\ = \sum_{(i,j) \in \mathcal{IJ}} \|\mathbf{e}_i^0 \cdot \mathbf{n}_i + \alpha_{ij} e_i^0\|^2_{L^2(\Gamma_{ij})} / (4\, \alpha_{ij}) \end{aligned}$$

We have therefore shown that the series $\sum_{\ell=1}^{\infty} \sum_{i \in \mathcal{I}_\mathcal{R}} \Big(\|(\Sigma_{a,i})^{\frac{1}{2}}\, e_i^\ell\|^2_{L^2(\mathcal{R}_i)} + \|(D_i)^{-\frac{1}{2}}\, \mathbf{e}_i^\ell\|^2_{L^2(\mathcal{R}_i)} \Big)$ converges. This means that the sequence $\left(\sum_{i \in \mathcal{I}_\mathcal{R}} \Big(\|(\Sigma_{a,i})^{\frac{1}{2}}\, e_i^\ell\|^2_{L^2(\mathcal{R}_i)} + \|(D_i)^{-\frac{1}{2}}\, \mathbf{e}_i^\ell\|^2_{L^2(\mathcal{R}_i)} \Big) \right)_{\ell \in \mathbb{N}}$ converges to zero, i.e. for $i \in \mathcal{I}_\mathcal{R}$, the sequence $(\mathbf{e}_i^\ell, e_i^\ell)_{\ell \in \mathbb{N}}$ tends to zero in $L^2(\mathcal{R}_i) \times \mathbf{L}^2(\mathcal{R}_i)$. Finally, using the relations (A.2.1)-(0) and (A.2.1)-(i), we obtain the convergence of the sequence $(\mathbf{e}_i^\ell, e_i^\ell)_{\ell \in \mathbb{N}}$ to 0 in $\mathbf{H}(\mathrm{div}, \mathcal{R}_i) \times V_i$.

A.3 Optimized Schwarz Method from algebraic point of view

Let $i \in \mathcal{I}$. We denote by $\Big((P_{i,m}^\ell)_{m \in \mathcal{I}_i^{f+}},\, (\Phi_{i,l}^\ell)_{l \in \mathcal{I}_i^T} \Big)$ the decomposition of $(\mathbf{p}_{i,h}^\ell, \phi_{i,h}^\ell)$ in their respective basis. We set: $P_i^\ell = (P_{i,m}^\ell)_{m \in \mathcal{I}_i^{f+}}$, $\Phi_i^\ell = (\Phi_{i,l}^\ell)_{l \in \mathcal{I}_i^T}$ and $S_i = (\int_{\widetilde{\mathcal{R}}_i} S_{f,i} \underline{\psi}_{i,l})_{l \in \mathcal{I}_i^T}$.

Let $\mathbb{M}^{i,i}_{\mathbf{p}} \in \mathbb{R}^{F_i^+ \times F_i^+}$ and $\mathbb{M}^{i}_{\phi} \in \mathbb{R}^{N_i \times N_i}$ be the symmetric, positive definite matrices such that:

$$\begin{aligned} &\forall\, (m,m') \in \mathcal{I}_i^{f+} \times \mathcal{I}_i^{f+} \quad (\mathbb{M}^{i,i}_{\mathbf{p}})_{m,m'} = \int_{\widetilde{\mathcal{R}}_i} (D_i)^{-1} \underline{\omega}_{i,m'} \cdot \underline{\omega}_{i,m}. \\ &\forall\, (l,l') \in \mathcal{I}_i \times \mathcal{I}_i \qquad\quad (\mathbb{M}^{i}_{\phi})_{l,l'} = \int_{\widetilde{\mathcal{R}}_i} \Sigma_{a,i} \underline{\psi}_{i,l}\, \underline{\psi}_{i,l'}. \end{aligned}$$

The matrices $\mathbb{M}^{i}_{\phi}$ are block diagonal. We consider $\widetilde{\mathbb{M}}^{i,i}_{\mathbf{p}} \in \mathbb{R}^{F_i^+ \times F_i^+}$ the symmetric matrix such that:

$$\forall j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}, \quad \begin{cases} \forall (m,m') \in \mathcal{I}_i^{f} \times \mathcal{I}_i^{f+} & (\widetilde{\mathbb{M}}^{i,i}_{\mathbf{p}})_{m,m'} = (\mathbb{M}^{i,i}_{\mathbf{p}})_{m,m'}, \\ \forall (m,m') \in \mathcal{I}_{i,ij}^{f} \times \mathcal{I}_i^{f+} & (\widetilde{\mathbb{M}}^{i,i}_{\mathbf{p}})_{m,m'} = (\mathbb{M}^{i,i}_{\mathbf{p}})_{m,m'} + \displaystyle\int_{\Gamma_{ij}} (\alpha^i_{ij})^{-1} \underline{\omega}_{i,m'} \cdot \mathbf{n}_i\, \underline{\omega}_{i,m} \cdot \mathbf{n}_i. \end{cases}$$

Let $j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}$. We call $\widetilde{\mathbb{M}}^{i,j}_{\mathbf{p}} \in \mathbb{R}^{F_i^+ \times F_j^+}$ the matrix linking $\mathbf{p}^{\ell}_{i,h}$ and $\mathbf{p}^{\ell}_{j,h}$:

$$\begin{aligned} &\forall (m,m') \in \mathcal{I}_i^{f} \times \mathcal{I}_j^{f+} \qquad (\widetilde{\mathbb{M}}^{i,j}_{\mathbf{p}})_{m,m'} = 0, \\ &\forall (m_i,m') \in \mathcal{I}_{i,ij}^{f} \times \mathcal{I}_j^{f+} \quad (\widetilde{\mathbb{M}}^{i,j}_{\mathbf{p}})_{m_i,m'} = \int_{\widetilde{\mathcal{R}}_j} (D_j)^{-1} \underline{\omega}_{j,m'} \cdot \underline{\omega}_{j,m_j} - \int_{\Gamma_{ij}} (\alpha^i_{ij})^{-1} \underline{\omega}_{j,m'} \cdot \mathbf{n}_j\, \underline{\omega}_{j,m_j} \cdot \mathbf{n}_j, \end{aligned}$$

where the expression of $(\widetilde{\mathbb{M}}^{i,j}_{\mathbf{p}})_{(m_i,m') \in \mathcal{I}_i^{ij,f} \times \mathcal{I}_j^{f+}}$ is obtained following the choice of the lifting (6.5.15) on the interface Γ_{ij}. Let $\mathbb{B}_{i,i} \in \mathbb{R}^{F_i^+ \times N_i}$ be the matrix linking the two unknowns inside the same subdomain:

$$\forall\, (m,l) \in \mathcal{I}_i^{f+} \times \mathcal{I}_i^{T} \quad (\mathbb{B}_{i,i})_{m,l} = \int_{\widetilde{\mathcal{R}}_i} \operatorname{div} \underline{\omega}_{i,m}\, \underline{\psi}_{i,l}.$$

Let $\widetilde{\mathbb{B}}_{i,j} \in \mathbb{R}^{F_i^+ \times N_j}$ be the matrix linking the two unknowns in neighbouring subdomains, cf. (6.5.15) for the relationship between m_i and m_j:

$$\forall\, (m,l) \in \mathcal{I}_i^{f+} \times \mathcal{I}_i \quad (\widetilde{\mathbb{B}}_{i,j})_{m,l} = \begin{cases} \displaystyle\int_{\widetilde{\mathcal{R}}_j} \operatorname{div} \underline{\omega}_{j,m_j}\, \underline{\psi}_{j,l} & \text{if} \quad m = m_i \in \mathcal{I}_{i,ij}^{f}, \\ 0 & \text{if} \quad m \in \mathcal{I}_i^{f}. \end{cases}$$

Equations (6.5.16)-(6.5.17) finally take the form of the following linear system:
For all $i \in \mathcal{I}$, find $(P_i^{\ell}, \Phi_i^{\ell})$ such that:

$$\begin{cases} \mathbb{M}^{i}_{\phi} \Phi_i^{\ell} + \mathbb{B}^{T}_{i,i} P_i^{\ell} &= S_i, \\ \mathbb{B}_{i,i} \Phi_i^{\ell} - \widetilde{\mathbb{M}}^{i,i}_{\mathbf{p}} P_i^{\ell} &= \displaystyle\sum_{j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}} \widetilde{\mathbb{M}}^{i,j}_{\mathbf{p}} P_j^{\ell-1} - \widetilde{\mathbb{B}}_{i,j} \Phi_j^{\ell-1}. \end{cases} \tag{A.3.1}$$

The discretization of algorithm (6.5.5) reads (the global matrix is not symmetric):

initialization: $\forall i \in \mathcal{I}$, $(P_i^0, \Phi_i^0) \in \mathbb{R}^{N_i} \times \mathbb{R}^{F_i^+}$ **is given.**
iterations: for $\ell = 1, \ldots,$ $\forall i \in \mathcal{I}$, **find** $(\Phi_i^{\ell}, P_i^{\ell}) \in \mathbb{R}^{N_i} \times \mathbb{R}^{F_i^+}$ **such that:**

$$\begin{pmatrix} -\widetilde{\mathbb{M}}^{i,i}_{\mathbf{p}} & \mathbb{B}_{i,i} \\ \mathbb{B}^{T}_{i,i} & \mathbb{M}^{i}_{\phi} \end{pmatrix} \begin{pmatrix} P_i^{\ell} \\ \Phi_i^{\ell} \end{pmatrix} = \begin{pmatrix} 0 \\ S_i \end{pmatrix} - \sum_{j \in \mathcal{I}_{\widetilde{\mathcal{R}}_i}} \begin{pmatrix} -\widetilde{\mathbb{M}}^{i,j}_{\mathbf{p}} & \widetilde{\mathbb{B}}_{i,j} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} P_j^{\ell-1} \\ \Phi_j^{\ell-1} \end{pmatrix}, \tag{A.3.2}$$

until convergence.

Algorithm (A.3.2) corresponds to a block Jacobi to solve (A.3.1). We can also use a GMRES algorithm. Algorithm (A.3.2) can be demonstrated using algebraic operations [A11, §6.2].

Appendix B

Fluid mechanics

B.1 The TrioCFD code

The TrioCFD code is often called a "finite volume element" code, since one can express the $\mathbf{P}^1_{nc} \times P^0$ scheme as a finite volume scheme. Moreover, when discretizing the Navier-Stokes equations, the convection operator is discretized using a finite volume schema, enabling stability.

B.1.1 Finite volume version of the $\mathbf{P}^1_{nc} \times P^0$ scheme

Let us first mention that the $\mathbf{P}^1_{nc} \times P^0$ scheme can be expressed as a finite volume scheme. Here, we denote by N_T be the number of simplices and N_F (resp. N_F^i), the number of faces in $\bar{\Omega}$ (resp. in $\bar{\Omega}$). We let $\mathcal{I}_D = \{1, \ldots, N_D\}$, where N_D is the dimension ($N_D = 2$ or 3). Recall that $\mathcal{I}_F = \{1, \ldots, N_F\}$ and $\mathcal{I}_T = \{1, \ldots, N_T\}$. We refer to §4.7 and §7.5 for the other notations.

For the discrete velocity, the unknowns are located on the barycenters of the faces. Let $F_f = K$, $f \in \mathcal{I}_F$ be a face. If $F_f \in \Omega$, it exists ℓ, ℓ' such that $F_f = K_\ell \cap K_{\ell'}$. The control volume related to F_f, denoted by ω_f is constructed joining the barycenters of K_ℓ and $K_{\ell'}$ to the vertices of F_f, as shown on Figure B.1 in $2D$. If $F_f \in \partial\Omega$, it exists ℓ such that $F_f \in \partial K_\ell$. The control volume ω_f is constructed joining the barycenters of K_ℓ to the vertices of F_f. For the discrete pressure, the unknowns are located on the barycenters of the elements and the related control volume is the element. We let $\mathbf{u}_h = (\mathbf{u}_f)_{f \in \mathcal{I}_F}$ and $p_h = (p_\ell)_{\ell \in \mathcal{I}_K}$.

After having substituted $(\mathbf{u}, p)$ by $(\mathbf{u}_h, p_h)$ in Equations (7.1.1), the weak formulation is obtained in two stages. First, we integrate Equation (7.1.1)-(i) on ω_f (resp. (7.1.1)-(ii) on K_ℓ), taking into account the discontinuities of the test functions. Next, we perform the scalar product between the N_T (resp. N_F^i) equations obtained and the vector representing a test function of the discrete pressure space (resp. of $\mathbf{X}_h$). We will also need these equalities:

$$\forall \ell \in \mathcal{I}_K, \forall f \in \mathcal{I}_{F,\ell}, \quad |\omega_f \cap K_\ell| = |K_\ell|/(d+1). \tag{B.1.1}$$

$$\forall \ell \in \mathcal{I}_K, \forall f \in \mathcal{I}_{F,\ell}, \quad \int_{\partial K_\ell} \phi_{f|K_\ell} \mathbf{n} = \mathcal{S}_{f,\ell}. \tag{B.1.2}$$

$$\forall f \in \mathcal{I}_F, \quad \int_{\partial \omega_f \cap K_\ell} \mathbf{n} = -\mathcal{S}_{f,\ell}, \tag{B.1.3}$$

$$\forall f \in \mathcal{I}_F, \quad \forall \phi \in L^2(\partial\omega_f): \quad \int_{\partial\omega_f} \phi = \sum_{\ell \in \mathcal{I}_{K,f}} \int_{\partial\omega_f \cap K_\ell} \phi. \tag{B.1.4}$$

B.1.2 Discrete operators

Using Equations (B.1.4) and (B.1.3), we define operators Δ_h and $\tilde{\nabla}_h$ such that:

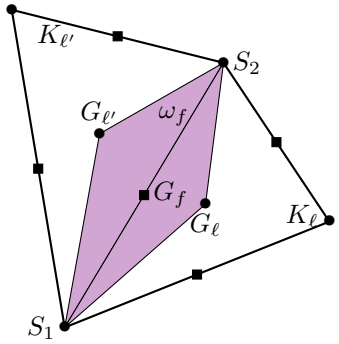


Figure B.1: $2D$ control volume ω_f related to F_f.

- Operator $\Delta_h : \mathcal{P}^1_{disc}(\mathcal{T}_h) \to L^2(\Omega)$, such that for all $v_h \in \mathcal{P}^1_{disc}(\mathcal{T}_h)$, for all $f \in \mathcal{I}^i_F$:

$$\begin{aligned} -\int_{\omega_f} \Delta_h v_h &= -\sum_{\ell \in \mathcal{I}_{K,f}} \int_{\partial\omega_f \cap K_\ell} \nabla(v_{h|K_\ell}) \cdot \mathbf{n} = -\sum_{\ell \in \mathcal{I}_{K,f}} \nabla(v_{h|K_\ell}) \cdot \int_{\partial\omega_f \cap K_\ell} \mathbf{n}, \\ &= \sum_{\ell \in \mathcal{I}_{K,f}} \nabla(v_{h|K_\ell}) \cdot \mathcal{S}_{f,\ell}. \end{aligned} \tag{B.1.5}$$

- Operator $\tilde{\nabla}_h : \mathcal{P}^0_{disc}(\mathcal{T}_h) \to \mathbf{L}^2(\Omega)$, such that for all $q_h \in \mathcal{P}^0_{disc}(\mathcal{T}_h)$, for all $f \in \mathcal{I}^i_F$:

$$\int_{\omega_f} \tilde{\nabla}_h q_h = \sum_{\ell \in \mathcal{I}_{K,f}} \int_{\partial\omega_f \cap K_\ell} q_{h|K_\ell}\, \mathbf{n} = \sum_{\ell \in \mathcal{I}_{K,f}} q_\ell \int_{\partial\omega_f \cap K_\ell} \mathbf{n} = -\sum_{\ell \in \mathcal{I}_{K,f}} q_\ell\, \mathcal{S}_{f,\ell}. \tag{B.1.6}$$

B.1.3 Discretization of the diffusion operator

Let $a_{fv}(\cdot,\cdot)$ be the continuous and coercive bilinear form representing the discretization of the diffusion term in Equation (7.1.1)-(i) using Equation (B.1.5).

$$\begin{aligned} a_{fv} : \mathbf{X}_{0,h} \times \mathbf{X}_{0,h} &\to \mathbb{R} \\ (\mathbf{v}_h, \mathbf{w}_h) &\mapsto -\nu \sum_{f \in \mathcal{I}^i_F} \mathbf{w}_f \cdot \int_{\omega_f} \Delta_h \mathbf{v}_h = \nu \sum_{f \in \mathcal{I}^i_F} \sum_{\ell \in \mathcal{I}_{K,f}} \nabla(\mathbf{v}_{h|K_\ell}) : \mathcal{S}_{f,\ell}. \end{aligned} \tag{B.1.7}$$

Thus, for all f, $f' \in \mathcal{I}_F$, for all d, $d' \in \mathcal{I}_D$, it holds:

$$a_{fv}(\phi_{f'}\, \mathbf{e}_d\, ,\, \phi_f\, \mathbf{e}_{d'}) = \delta_{d,d'}\, \nu\, A_{f,f'} \tag{B.1.8}$$

where :

$$A_{f,f'} = \begin{cases} |K_\ell|^{-1}\, \mathcal{S}_{f',\ell} \cdot \mathcal{S}_{f,\ell} & \text{if } f \neq f' \text{ and } F_f,\, F_{f'} \subset \partial K_\ell \\ \left(\displaystyle\sum_{\ell \in \mathcal{I}_{K,f}} |K_\ell|^{-1} \right) |F_f|^2 & \text{if } f = f' \end{cases} . \tag{B.1.9}$$

The bilinear form $a_h(\cdot,\cdot)$ is the same as that of the finite element method:

$$a_{fv}(\phi_f\, \mathbf{e}_d\, ,\, \phi_{f'}\, \mathbf{e}_{d'}) = \delta_{d,d'}\, \nu\, (\phi_f, \phi_{f'})_h.$$

B.1.4 Discretization of the source term

Suppose that $\mathbf{f} \in \mathbf{L}^2(\Omega)$. The discretization of the source term in Equation (7.1.1)-(i) reads:

$$\begin{aligned} \ell_{\mathbf{f}} : \mathbf{X}_{0,h} &\to \mathbb{R} \\ \mathbf{v}_h &\mapsto (\mathbf{f}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} = \sum_{f \in \mathcal{I}_F^i} \left(\mathbf{v}_f \cdot \int_{\omega_f} \mathbf{f}\, \phi_f \right) . \end{aligned} \tag{B.1.10}$$

In the finite volume method, we should use instead of $\ell_{\mathbf{f}}^{fv}(\cdot)$:

$$\ell_{\mathbf{f}}^{fv}(\mathbf{v}_h) = \sum_{f \in \mathcal{I}_F^i} \left(\mathbf{v}_f \cdot \int_{\omega_f} \mathbf{f} \right). \tag{B.1.11}$$

Suppose $\mathbf{f} \in \mathbf{L}^\infty(\Omega)$. Using Cauchy-Schwarz inequality, we have:

$$\begin{aligned} \left| \sum_{\ell \in \mathcal{I}_{K,f}} \int_{K_\ell} \mathbf{f}\, \phi_f - \int_{\omega_f} \mathbf{f} \right| &= \sum_{\ell \in \mathcal{I}_{K,f}} \left| \left(\sum_{f' \in \mathcal{I}_{F,\ell}, f' \neq f} \int_{\omega_{f'} \cap K_\ell} \mathbf{f}\, \phi_f - N_D \int_{\omega_f \cap K_\ell} \mathbf{f}\, \lambda_{f,\ell} \right) \right| \\ &\leq \sum_{\ell \in \mathcal{I}_{K,f}} \left(\sum_{f' \in \mathcal{I}_{F,\ell}, f' \neq f} \left(|\omega_{f'} \cap K_\ell|^{\frac{1}{2}} + N_D \, \|\lambda_{f,\ell}\|_{L^2(\omega_f \cap K_\ell)} \right) \|\mathbf{f}\|_{\mathbf{L}^2(K_\ell)} \right) \end{aligned}$$

We have: $\|\lambda_{f,\ell}\|^2_{L^2(\omega_f \cap K_\ell)} \leq \|\lambda_{f,\ell}\|^2_{L^2(K_\ell)} = |K_\ell| \, \dfrac{N_D!\, 2!}{(N_D+2)!}$. We obtain, using (B.1.1):

$$\left| \sum_{\ell \in \mathcal{I}_{K,f}} \int_{K_\ell} \mathbf{f}\, \phi_f - \int_{\omega_f} \mathbf{f}_h \right| \leq \left(\frac{|K_\ell|}{N_D + 1} \right)^{\frac{1}{2}} \|\mathbf{f}\|_{\mathbf{L}^2(K_\ell)} \lesssim h^{N_D} \, \|\mathbf{f}\|_{\mathbf{L}^\infty(K_\ell)}. \tag{B.1.12}$$

The bilinear form obtained with the finite volume method is an approximation of that obtained with the finite element method. Using it may deteriorate the accuracy of the $\mathbf{P}^1_{nc} \times P^0$ and $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ schemes.

B.1.5 Discretization of the gradient of the pressure

Let $b_{fv}(\cdot, \cdot)$ be the continuous bilinear form representing the discretization of the gradient term for discrete pressures P^0 in Equation (7.1.1)-(i) (cf. (B.1.6)).

$$\begin{aligned} b_{fv} : \mathbf{X}_{0,h} \times \mathcal{P}^0_{disc}(\mathcal{T}_h) &\to \mathbb{R} \\ (\mathbf{v}_h, q_h) &\mapsto \sum_{f \in \mathcal{I}_F^i} \mathbf{v}_f \cdot \int_{\omega_f} \tilde{\nabla}_h q_h = - \sum_{f \in \mathcal{I}_F^i} \sum_{\ell \in \mathcal{I}_{K,f}} q_\ell \, \mathbf{v}_f \cdot \mathcal{S}_{f,\ell} \ . \end{aligned} \tag{B.1.13}$$

Thus, for all $d \in \mathcal{I}_D$, for all $f \in \mathcal{I}_F$, for all $\ell \in \mathcal{I}_K$, it holds:

$$b_{fv}(\phi_f \, \mathbf{e}_d, \psi_\ell) = (B_{0,d})_{f,\ell} := \begin{cases} -\mathcal{S}_{f,\ell} \cdot \mathbf{e}_d & \text{if } F_f \subset \partial K_\ell \\ 0 & \text{else} \end{cases} . \tag{B.1.14}$$

Let $\tilde{b}_{fv}(\cdot, \cdot)$ be the bilinear form discretizing Equation (7.1.1)-(ii):

$$\begin{aligned} \tilde{b}_{fv} : \mathbf{X}_{0,h} \times \mathcal{P}^0_{disc}(\mathcal{T}_h) &\to \mathbb{R} \\ (\mathbf{v}_h, q_h) &\mapsto \sum_{\ell \in \mathcal{I}_K} q_\ell \int_{\partial K_\ell} \mathbf{v}_h \cdot \mathbf{n} \ . \end{aligned}$$

We have then:

$$\tilde{b}_{fv}(\mathbf{v}_h, q_h) = \sum_{\ell \in \mathcal{I}_K} q_\ell \int_{K_\ell} \operatorname{div} \mathbf{v}_{h|K_\ell} = \sum_{\ell \in \mathcal{I}_K} q_\ell \sum_{f \in \mathcal{I}_{F,\ell}} \mathbf{v}_f \cdot \int_{\partial K_\ell} \phi_{f|K_\ell} \mathbf{n}.$$

Using Equation (B.1.2), it holds:

$$\forall (d, \ell, f) \in \mathcal{I}_D \times \mathcal{I}_K \times \mathcal{I}_F, \quad \tilde{b}_{fv}(\phi_f \, \mathbf{e}_d, \psi_\ell) = -b_{fv}(\phi_f \, \mathbf{e}_d, \psi_\ell). \tag{B.1.15}$$

We use only the bilinear form $b_{0,h}(\cdot,\cdot)$, which is the same as that of the coupling term between $\mathbf{P}^1_{nc}$ velocity and P^0 pressure in the finite element method: $b_{fv}(\phi_f \, \mathbf{e}_d, \psi_\ell) = -(\operatorname{div}(\phi_f \, \mathbf{e}_d), 1)_{L^2(K_\ell)}$.

Note that a finite volume approach is proposed in [175] to describe the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ scheme. For the P^1 component of the pressure, the control volume of a vertex corresponds to the P^2 submesh surrounding the vertex.

B.1.6 Linear system

Recall that $\mathbf{u}_h = \sum_{f \in \mathcal{I}^i_F} \mathbf{u}_f \phi_f$. We let $U_{h,d} = (\mathbf{u}_f \cdot \mathbf{e}_d)_{f \in \mathcal{I}^i_F}$ and $U_h = (U_{h,d})_{d \in \mathcal{I}_D}$.

Recall that $p_{0,h} = \sum_{\ell \in \mathcal{I}_K} p_\ell \psi_\ell$. We let $\underline{P}_{0,h} = (p_\ell)_{\ell \in \mathcal{I}_K}$. We have $p_{1,h} = \sum_{i \in \mathcal{I}_S} p_i \lambda_i$, and we let $\underline{P}_{1,h} = (\underline{P}_{1,h})_{i \in \mathcal{I}_S}$.

Let $A \in \mathbb{R}^{N^i_F} \times \mathbb{R}^{N^i_F}$ be the stiffness matrix for scalar functions in $X_{0,h}$, where $A_{f,f'}$ is defined in Equation (B.1.9). Let $[A] \in \mathbb{R}^{(N_D \times N^i_F)} \times \mathbb{R}^{(N_D \times N^i_F)}$ be the stiffness matrix for vectorial functions of $\mathbf{X}_{0,h}$, such that: for $d, d' \in \mathcal{I}_D$, $[A]_{d,d'} = \delta_{d,d'} A$ (i.e. $[A]$ is block diagonal).

Let $\underline{B}_{0,d} \in \mathbb{R}^{N^i_F} \times \mathbb{R}^{N_K}$ where $(\underline{B}_{0,d})_{f,\ell}$ is given by (B.1.14). We set $[\underline{B}_0] := \left(\underline{B}_{0,d}\right)_{d \in \mathcal{I}_D}$.

Let $\underline{B}_{1,d} \in \mathbb{R}^{N^i_F} \times \mathbb{R}^{N_S}$ be such that $(\underline{B}_{1,d})_{f,i} := (\phi_f \, \mathbf{e}_d, \nabla \psi_i)_{\mathbf{L}^2(\Omega)}$. We set $[\underline{B}_1] := \left(\underline{B}_{1,d}\right)_{d \in \mathcal{I}_D}$.

Let $\underline{F}_{h,d} \in \mathbb{R}^{N^i_F}$ be such that $\underline{F}_{h,d} = \left(\ell_{\mathbf{f}}(\phi_f \, \mathbf{e}_d)_{f \in \mathcal{I}^i_F}\right)$, and we set $\underline{F}_h = \left(\underline{F}_{h,d}\right)_{d \in \mathcal{I}_D}$.

$$\left\{ \begin{array}{l} \text{Find } (U_h, \underline{P}_{0,h}, \underline{P}_{1,h}) \text{ such that} \\ \nu \, [A] \, U_h + [\underline{B}_0] \, \underline{P}_{0,h} + [\underline{B}_1] \, \underline{P}_{1,h} = \underline{F}_h \\ \qquad -[\underline{B}_0]^T \, U_h - [\underline{B}_1]^T \, U_h = 0, \\ \qquad \qquad \displaystyle\sum_{\ell \in \mathcal{I}_K} |K_\ell| \, p_\ell = 0, \\ \qquad \displaystyle\sum_{\ell \in \mathcal{I}_K} |K_\ell| \sum_{i \in \mathcal{I}_{S,\ell}} p_i = 0. \end{array} \right. \tag{B.1.16}$$

The matrix $[A]$ is a sparse, symmetric, positive-definite matrix. The matrix $[\underline{B}_0]$ is a sparse matrix of rank $N_T - 1$. In order to compute $p_{0,h} \in Q_{0,h}$, we impose p_{ℓ_0} for some $\ell_0 \in \mathcal{I}_K$, cancelling column ℓ_0 from matrix $[\underline{B}_0]$ and line ℓ_0 from vector $\underline{P}_{0,h}$, and transferring $[\underline{B}_0]_{:,\ell_0} p_{\ell_0}$ to the right-hand-side of the first equation of (B.1.16). The matrix $[\underline{B}_1]$ is a sparse matrix of rank $N_S - 1$. In order to compute $p_{1,h} \in Q_{1,h}$, we impose p_{i_0} for some $i_0 \in \mathcal{I}_S$, cancelling column i_0 from matrix $[\underline{B}_1]$ and line i_0 from vector $\underline{P}_{1,h}$, and transferring the vector $[\underline{B}_1]_{:,i_0} \, p_{i_0}$ to the right-hand-side of the first equation of (B.1.16). Let $[B_1] = [\underline{B}_1]_{:,i \in \mathcal{I}_S \setminus i_0}$. Problem (B.1.16) reads:

$$\left\{ \begin{array}{l} \text{Find } (U_h, P_{0,h}, P_{1,h}) \text{ such that} \\ \nu \, [A] \, U_h + [B_0] \, P_{0,h} + [B_1] \, P_{1,h} = F_h := \underline{F}_h - [\underline{B}_0]_{:,\ell_0} \, p_{\ell_0} - [\underline{B}_1]_{:,i_0} \, p_{i_0} \\ \qquad -[B_0]^T \, U_h - [B_1]^T \, U_h = 0 \end{array} \right. , \tag{B.1.17}$$

where $[B_0]$ and $[B_1]$ are of maximal rank. The matrix $[B] = ([B_0], [B_1])$ is also of maximal rank. Let $P_h := (P_{0,h}, P_{1,h})^T$. Problem (B.1.17) reads:

$$\left\{ \begin{array}{l} \text{Find } (U_h, P_h) \text{ such that:} \\ \nu\,[A]\,U_h + [B]\,P_h \;=\; F_h \\ \\ \qquad -[B]^T\,U_h \;=\; 0 \end{array} \right. . \tag{B.1.18}$$

The computation of $\|p_h\|^2_{L^2(\Omega)}$ reads:

$$\|p_h\|^2_{L^2(\Omega)} = \sum_{\ell\in\mathcal{I}_K} |K_\ell|\,(p_\ell)^2 + \widetilde{C}_D \sum_{\ell\in\mathcal{I}_K} |K_\ell| \left(\sum_{i\in\mathcal{I}_{S,\ell}} (p_i)^2 + \sum_{i,j\in\mathcal{I}_{S,\ell}\; i\neq j} p_i\,p_j \right) + C_D \sum_{\ell\in\mathcal{I}_K} |K_\ell|\,p_\ell \sum_{i\in\mathcal{I}_{S,\ell}} p_i, \tag{B.1.19}$$

where we have: $\widetilde{C}_D = 1/6$, $C_D = 2/3$ if $N_D = 2$ and $\widetilde{C}_D = 1/10$, $C_D = 1/2$ if $N_D = 3$.

B.1.7 Discretization of the convection

Let $\tilde{c}_h(\cdot,\cdot,\cdot)$ be the trilinear form discretizing the convective term in (7.1.1)-(i):

$$\begin{array}{rcl} \tilde{c}_h : \mathbf{X}_{0,h}\times\mathbf{X}_{0,h}\times\mathbf{X}_{0,h} & \to & \mathbb{R} \\ (\mathbf{u}_h;\mathbf{v}_h,\mathbf{w}_h) & \mapsto & \displaystyle\sum_{f\in\mathcal{I}_F^i} \left(\mathbf{w}_f \cdot \int_{\omega_f} (\mathbf{v}_h\cdot\nabla_h)\mathbf{u}_h \right) = \sum_{f\in\mathcal{I}_F^i} \sum_{\ell\in\mathcal{I}_{K,f}} \left(\mathbf{w}_f \cdot \int_{\omega_f\cap K_\ell} (\mathbf{v}_h\cdot\nabla)\mathbf{u}_h \right) . \end{array}$$

Notice that for $\mathbf{u}_h \in \mathbf{V}_{0,h}$, it stands:

$$\tilde{c}_h(\mathbf{u}_h;\mathbf{v}_h,\mathbf{w}_h) = \sum_{f\in\mathcal{I}_F^i} \sum_{\ell\in\mathcal{I}_{K,f}} \left(\mathbf{w}_f \cdot \int_{\partial\omega_f\cap K_\ell} \mathbf{u}_h\,(\mathbf{v}_h\cdot\mathbf{n}) \right).$$

Using this trilinear form, one cannot prove stability. FOr this reason, in the TrioCFD code, a finite volume scheme is used to discretize the convective term. We preset here numerical results computed with an upwind scheme, described below, and a MUSCL scheme[1] with a Van Leer slope limiter [189].

Let $F_{f,f'} = F_{f',f} := \partial\omega_f \cap \partial\omega_{f'}$ and $\mathcal{S}_{f,f'} = -\mathcal{S}_{f',f}$ be the normal vector leaving $\partial\omega_f$ on $F_{f,f'}$ and such that $|\mathcal{S}_{f,f'}| = |F_{f,f'}|$ (see Figure B.2 for $N_D = 2$). Let $G_{f,f'}$ be the barycenter of facet $F_{f,f'}$.

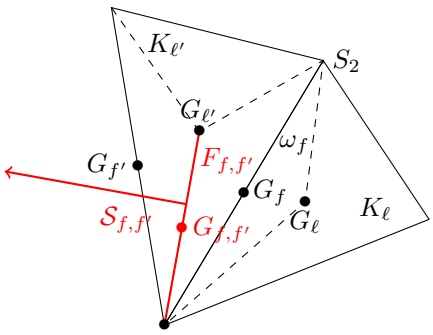


Figure B.2: Facet $F_{f,f'}$ and outgoing normal vector $\mathcal{S}_{f,f'}$.

The trilinear form related to the upwind scheme is such that:

$$\begin{array}{rcl} c_h : \mathbf{X}_{0,h}\times\mathbf{X}_{0,h}\times\mathbf{X}_{0,h} & \to & \mathbb{R} \\ (\mathbf{u}_h;\mathbf{v}_h,\mathbf{w}_h) & \mapsto & \displaystyle\sum_{f\in\mathcal{I}_F^i} \mathbf{w}_f \cdot \sum_{f'\in\mathcal{I}_{F,f}} \mathbf{u}_{f,f'}\,(\mathbf{v}_h(G_{f,f'})\cdot\mathcal{S}_{f,f'}), \end{array}$$

[1] MUSCL means Monotonic Upstream-centered Scheme for Conservation Laws.

where $\mathcal{I}_{F,f} = \{f' \in \mathcal{I}_F,\ f' \neq f \,|\, \partial F_f \cap \partial F_{f'} \neq \emptyset\}$ and $\mathbf{u}_{f,f'}$ is defined as follows

$$\mathbf{u}_{f,f'} = \begin{cases} \mathbf{u}_f & \text{if} \quad \mathbf{v}_h(G_{f,f'}) \cdot \mathcal{S}_{f,f'} > 0 \text{ on } F_{f,f'} \\ \mathbf{u}_{f'} & \text{else} \end{cases} . \tag{B.1.20}$$

Let $f \in \mathcal{I}_F^i$. Using the MUSCL scheme [189] in $2D$, the flow across the edges of $\partial\omega_f$, $(F_{f,f'})_{f' \in \mathcal{I}_{F,f}}$, is approximated with the following scheme:

$$\int_{F_{f,f'}} \mathbf{v}_h(\mathbf{v}_h \cdot \mathbf{n}) \approx \frac{1}{6} \left(\mathbf{v}_{h|K_{\ell'}}(S_1)\mathbf{V}_1 \cdot \mathcal{S}_{f,f'} + 4\, \mathbf{v}_h(G_{f,f'})\mathbf{V}_{f,f'} \cdot \mathcal{S}_{f,f'} + \mathbf{v}_h(G_{\ell'})\mathbf{v}_h(G_{\ell'}) \cdot \mathcal{S}_{f,f'} \right),$$

where the notations are shown on Figure B.3 and where $\mathbf{V}_1$ and $\mathbf{V}_{f,f'}$ are such that:

$$\text{If } \mathbf{v}_h(G_{f,f'}) \cdot \mathcal{S}_{f,f'} > 0 : \quad \begin{cases} \mathbf{V}_1 &=& \mathbf{v}_f + \nabla\tilde{\mathbf{v}}_f : \overrightarrow{G_f S_1} \\ \mathbf{V}_{f,f'} &=& \mathbf{v}_f + \nabla\tilde{\mathbf{v}}_f : \overrightarrow{G_f G_{f,f'}} \end{cases} \quad \text{else:} \quad \begin{cases} \mathbf{V}_1 &=& \mathbf{v}_{f'} + \nabla\tilde{\mathbf{v}}_{f'} : \overrightarrow{G_{f'} S_1} \\ \mathbf{V}_{f,f'} &=& \mathbf{v}_{f'} + \nabla\tilde{\mathbf{v}}_{f'} : \overrightarrow{G_{f'} G_{f,f'}} \end{cases} .$$

We use the Van Leer slope limiter in order to calculate $\nabla\tilde{\mathbf{v}}_f$: Let $F_f = S_1 S_2 = \partial K_\ell \cap \partial K_{\ell'}$. We have:

$$\forall\, d,\, d' \in \mathcal{I}_D,\ (\nabla\tilde{\mathbf{v}}_f)_{d,d'} = \text{Van Leer}\left((\nabla\mathbf{v}_{h|K_\ell})_{d,d'}, (\nabla\mathbf{v}_{h|K_{\ell'}})_{d,d'} \right)$$

$$\text{with, for all } a,\, b \in \mathbb{R}\text{: Van Leer}(a,b) = \begin{cases} \dfrac{2\,a\,b}{a+b} & \text{if } a\,b > 0 \\ 0 & \text{else} \end{cases} .$$

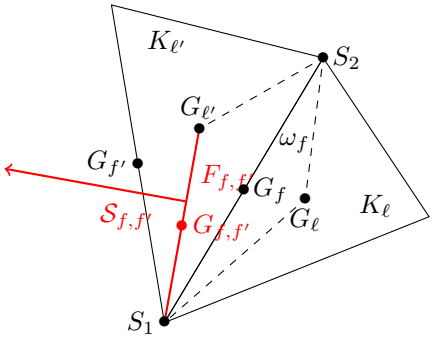


Figure B.3: Control volume ω_f associated to G_f

B.1.8 Time discretization

Let $T > 0$ be the final simulation time. The semi-discretization in space of Equations (7.1.1), with $\mathbf{f} \in \mathbf{L}^2(\Omega)$, formulated with the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ scheme reads:

$$\begin{cases} \text{For all } t \in (0,T), \text{ find } (\mathbf{u}_h(t), p_h(t)) \in \mathbf{X}_{0,h} \times Q_h \text{ such that:} \\ \\ \quad \mathbf{u}_h(0) &=& \Pi_{cr}(\mathbf{u}_0) \\ \dfrac{d}{dt}(\mathbf{u}_h, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} + \nu(\mathbf{u}_h, \mathbf{v}_h)_h + c_h(\mathbf{u}_h, \mathbf{u}_h, \mathbf{v}_h) + b_h(\mathbf{v}_h, p_h) &=& (\mathbf{f}, \mathbf{v}_h)_{\mathbf{L}^2(\Omega)} \quad \forall \mathbf{v}_h \in \mathbf{X}_{0,h} \\ -b_h(\mathbf{u}_h, q_h) &=& 0 \qquad\qquad\qquad\quad \forall q_h \in Q_h \end{cases} , \tag{B.1.21}$$

We call $M \in \mathbb{R}^{N_F^i} \times \mathbb{R}^{N_F^i}$ the mass matrix of the velocity components, which is a diagonal matrix: for all $f,\, f' \in \mathcal{I}_F^i$, $M_{f,f'} = \delta_{f,f'}|\omega_f|$. Let $M_{cr} \in \mathbb{R}^{N_F^i} \times \mathbb{R}^{N_F^i}$ be the mass matrix of $X_{0,h}$ functions, such that: for all $f,\, f' \in \mathcal{I}_F^i$, $M_{f,f'} = \int_\Omega \phi_f\, \phi_{f'}$. When $N_D = 2$, $M_{cr} = M$. When $N_D = 3$, $M_{cr} \neq M$,

but one can prove that these matrices are equivalent. Let $[M]$ the block diagonal matrix such that: $[M]_{d,d'} = \delta_{d,d'} M$. Let $[A] \in \mathbb{R}^{(N_D \times N_F^i)} \times \mathbb{R}^{(N_D \times N_F^i)}$ be the stiffness matrix for vectorial functions of $\mathbf{X}_{0,h}$. Let $[B] \in \mathbb{R}^{N_P} \times \mathbb{R}^{(N_D \times N_F^i)}$ The nonlinear system related to (B.1.21) reads:

$$\left\{ \begin{array}{l} \text{For all } t \in (0,T), \text{ find } (U_h(t), P_h(t)) \text{ such that:} \\ \qquad\qquad\qquad\qquad U_h(0) = U_0 \\ [M]\dfrac{d}{dt}U_h + \nu\,[A]\,U_h + [L(U_h)]U_h + [B]\,P_h = F_h(t) \\ \qquad\qquad\qquad\qquad -[B]^T\,U_h = 0 \end{array} \right. , \tag{B.1.22}$$

where $U_0 \in \mathbb{R}^{(N_D \times N_F^i)}$ is such that $U_0 = (U_{0,d})_{d \in \mathcal{I}_D}$ and $U_{0,d} = (\pi_f(\mathbf{u}_0 \cdot \mathbf{e}_d))_{f \in \mathcal{I}_F^i}$.

In the TrioCFD code, numerical simulations of Stokes problem are obtained as the asymptotic limit of a transient state. There exist different time discretization schemes [10]. In the present work, all the simulations are done using a semi-implicit scheme. Let δt_n be the time step at step n, and U_h^n be the approximation of $U_h(t_n)$. We approximate $[L(U_h(t_n))]U_h(t_n)$ by $[L(U_h^{n-1})]U_h^n$ and we use Euler's backwards scheme. Let $U_h^0 = U_0$. Compute δt_1. The space-time discretization scheme for (7.1.1) reads:

$$\left\{ \begin{array}{l} \text{Find } (U_h^n, P_h^n) \text{ such that for all } n \in \mathbb{N}^\star \\ \dfrac{1}{\delta t_n}[M](U_h^n - U_h^{n-1}) + \nu\,[A]U_h^n + [L(U_h^{n-1})]U_h^n + [B]P_h^n = F_h^n \\ \qquad\qquad\qquad\qquad -[B]^T U_h^n = 0 \end{array} \right. . \tag{B.1.23}$$

The time step δt_{n+1} is evaluated at each iteration. We stop the iteration when $\sum_{n'=1}^{n-1} \delta t_{n'} \geq T$. Let us set $[K(U_h^{n-1})] := \dfrac{1}{\delta t_n}[M] + \nu\,[A] + [L(U_h^{n-1})]$ and $\tilde{F}_h^n = F_h^n + \dfrac{1}{\delta t_n}[M]\,U_h^{n-1}$.

The time-discretization scheme (B.1.23) reads:

$$\left\{ \begin{array}{l} \text{Find } (U_h^n, P_h^n) \text{ such that } U_h^0 = U_0 \text{ and for all } n \in \mathbb{N}^\star \\ [K(U_h^{n-1})]U_h^n + [B]P_h^n = \tilde{F}_h^n \\ \qquad\qquad -[B]^T U_h^n = 0 \end{array} \right. . \tag{B.1.24}$$

The matrix $[K(U_h^{n-1})]$ is a sparse positive-definite matrix. For a given time step $n \in \mathbb{N}^\star$, the linear system (B.1.24) is not solved exactly. Instead, one relies on the following three steps scheme [89, 240]:

$$\begin{array}{ll} \text{Prediction:} & \text{Solve in } U_h^\star : [K(U_h^{n-1})]U_h^\star = \tilde{F}_h^n - [B]P_h^{n-1}. \\ \text{Pressure solver:} & \text{Solve in } \delta P_h : [B]^T[M]^{-1}[B]\,\delta P_h = \dfrac{1}{\delta t_n}[B]^T U_h^\star. \\ \text{Update:} & U_h^n = U_h^\star - \dfrac{1}{\delta t_n}[M]^{-1}[B]\,\delta P_h, \quad P_h^n = P_h^{n-1} + \delta\,P_h. \end{array} \tag{B.1.25}$$

In practice, U_0 is evaluated from the user data and P_h^0 solves $[B]^T[M]^{-1}[B]\,P_h^0 = [B]^T[M]^{-1}\,F_h^0$. Notice that this choice will not affect (U_h^n, P_h^n). The diagonal matrix $[M]$ is of rank $N_D \times N_F^i$. The sparse matrix $[B]$ is of maximal rank. Hence, the matrix of the second step $[B]^T[M]^{-1}[B]$ is a sparse, symmetric and positive-definite matrix. We eventually can build it explicitly and store it. Moreover, the third step is straightforward.

If we were using an explicit time discretization scheme (i.e. writing $\nu[A]U_h^{n-1} + [L(U_h^{n-1})]U_h^{n-1}$ instead of $\nu[A]U_h^n + [L(U_h^{n-1})]U_h^n$ in Equation (B.1.23)), we would impose a Courant-Friedrich-Levy (CFL) condition on the time step $\tilde{\delta} t_n$ so that: $\tilde{\delta} t_n = \min(f\,\delta t_{n,\text{stab}}\,,\,\delta t_{\max})$, where:

- The coefficient f and the time step $\delta t_{\max}$ are set by the user.
- The time step $\delta t_{n,\text{stab}}$ is computed at each time step $n \in \mathbb{N}$ this:

$$\frac{1}{\delta t_{n,\text{stab}}} = \frac{1}{\delta t_{n,\text{conv}}} + \frac{1}{\delta t_{\text{diff}}} \quad \text{with} \quad \delta t_{n,\text{conv}} = \frac{h}{\max\limits_{d \in \mathcal{I}_D,\, f \in \mathcal{I}_F^i} (|(U_h^{n-1})_{d,f}|)} \quad \text{and} \quad \delta t_{\text{diff}} = \frac{h^2}{2\,\nu}. \tag{B.1.26}$$

In most cases, $\delta t_{\text{diff}} < \delta t_{n,\text{conv}}$. Using algorithm (B.1.25), we usually set $\delta t_n = \min(f\,\delta t_{n,\text{conv}}, \delta t_{\max})$, which reduces efficiently the simulation time compare to using an explicit time discretization algorithm.

In the present work, the solver used for the first step of algorithm (B.1.25) is the iterative solver GMRES from the PETSc library [222]. The solver used for the second step of algorithm (B.1.25) is based on PETSc solvers using either a direct method (Cholesky decomposition) or iterative methods (preconditioned conjugate gradient method with SSOR preconditioner).

B.2 Proof of Theorem 7.5.3

Let $(\mathbf{u}_h, p_h) \in \mathcal{X}_h$ be the solution to Problem (7.5.32) with $p_h = p_{0,h} + p_{1,h}$, $(p_{0,h}, p_{1,h}) \in Q_{0,h} \times Q_{1,h}$. We prove (7.5.45) and (7.5.46), following the proofs of [121, Theorems 3 and 4]. For all $\mathbf{v}_h \in \mathbf{V}_{cr}$:

$$\|\mathbf{u} - \mathbf{u}_h\|_h \leq \|\mathbf{u} - \mathbf{v}_h\|_h + \|\mathbf{v}_h - \mathbf{u}_h\|_h, \tag{B.2.1}$$

$$\|\mathbf{v}_h - \mathbf{u}_h\|_h^2 = -(\mathbf{u} - \mathbf{v}_h, \mathbf{v}_h - \mathbf{u}_h)_h + (\mathbf{u} - \mathbf{u}_h, \mathbf{v}_h - \mathbf{u}_h)_h \tag{B.2.2}$$

Let $\mathbf{v}_h \neq \mathbf{u}_h$. Using Cauchy-Schwarz in (B.2.2), it holds:

$$\|\mathbf{v}_h - \mathbf{u}_h\|_h^2 \leq \|\mathbf{u} - \mathbf{v}_h\|_h \, \|\mathbf{v}_h - \mathbf{u}_h\|_h + \frac{(\mathbf{u} - \mathbf{u}_h, \mathbf{v}_h - \mathbf{u}_h)_h}{\|\mathbf{v}_h - \mathbf{u}_h\|_h} \|\mathbf{v}_h - \mathbf{u}_h\|_h.$$

We obtain that

$$\|\mathbf{v}_h - \mathbf{u}_h\|_h \leq \|\mathbf{u} - \mathbf{v}_h\|_h + \sup_{\mathbf{w}_h \in \mathbf{V}_{cr}\setminus\{0\}} \frac{|(\mathbf{u} - \mathbf{u}_h, \mathbf{w}_h)_h|}{\|\mathbf{w}_h\|_h}. \tag{B.2.3}$$

We deduce from (B.2.1) and (B.2.3) that for all $\mathbf{v}_h \in \mathbf{V}_{cr}$:

$$\|\mathbf{u} - \mathbf{u}_h\|_h \leq 2\|\mathbf{u} - \mathbf{v}_h\|_h + \sup_{\mathbf{w}_h \in \mathbf{V}_{cr}\setminus\{0\}} \frac{|(\mathbf{u} - \mathbf{u}_h, \mathbf{w}_h)_h|}{\|\mathbf{w}_h\|_h}.$$

Hence:

$$\|\mathbf{u} - \mathbf{u}_h\|_h \leq 2 \inf_{\mathbf{w}_h \in \mathbf{V}_{cr}} \|\mathbf{u} - \mathbf{w}_h\|_h + \sup_{\mathbf{w}_h \in \mathbf{V}_{cr}\setminus\{0\}} \frac{|(\mathbf{u} - \mathbf{u}_h, \mathbf{w}_h)_h|}{\|\mathbf{w}_h\|_h}. \tag{B.2.4}$$

To bound the first term in the right-hand side of (B.2.4), we need the Fortin operator. Using (7.5.38), we have indeed:

$$\inf_{\mathbf{w}_h \in \mathbf{V}_{cr}} \|\mathbf{u} - \mathbf{w}_h\|_h \leq \|\mathbf{u} - \Pi_h \mathbf{u}\|_h \lesssim \sigma^{2d}\, h |\mathbf{u}|_{\mathbf{H}^2(\Omega)}. \tag{B.2.5}$$

Let us bound the second term in the right-hand side of (B.2.4). Consider $\mathbf{w}_h \in \mathbf{V}_{cr}$. It holds: $\nu(\mathbf{u}_h, \mathbf{w}_h)_h = (\mathbf{f}, \mathbf{w}_h)_{\mathbf{L}^2(\Omega)}$. Since $\mathbf{f} = -\nu\Delta\mathbf{u} + \nabla p$, we obtain by integration by parts:

$$\nu(\mathbf{u}_h, \mathbf{w}_h)_h = \nu(\mathbf{u}, \mathbf{w}_h)_h - \nu \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\boldsymbol{\nabla}\,\mathbf{u} : \mathbf{n}_{f,\ell}, \mathbf{w}_h)_{\mathbf{L}^2(F_f)} + \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (p, \mathbf{w}_h \cdot \mathbf{n}_{f,\ell})_{L^2(F_f)}.$$

Hence, for all $\mathbf{w}_h \in \mathbf{V}_{cr}$:

$$(\mathbf{u} - \mathbf{u}_h, \mathbf{w}_h)_h = \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (\boldsymbol{\nabla}\,\mathbf{u} : \mathbf{n}_{f,\ell}, \mathbf{w}_h)_{\mathbf{L}^2(F_f)} - \nu^{-1} \sum_{\ell \in \mathcal{I}_K} \sum_{f \in \mathcal{I}_{F,\ell}} (p, \mathbf{w}_h \cdot \mathbf{n}_{f,\ell})_{L^2(F_f)}. \tag{B.2.6}$$

We apply (7.4.2) to all the components of $\boldsymbol{\nabla}\,\mathbf{u} \in \mathbb{H}^1(\Omega)$ and $\mathbf{w}_h \in \mathbf{X}_{0,h}$ to obtain:

$$|(\mathbf{u} - \mathbf{u}_h, \mathbf{w}_h)_h| \lesssim \sigma h \left(|\mathbf{u}|_{\mathbf{H}^2(\Omega)} + \nu^{-1} |p|_{H^1(\Omega)} \right) \|\mathbf{w}_h\|_h. \tag{B.2.7}$$

Summing (B.2.5) and (B.2.7), we can bound (B.2.2). We deduce (7.5.45).

Let us prove (7.5.46). It holds:

$$\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{L}^2(\Omega)} = \sup_{\mathbf{w}\setminus\{0\} \in \mathbf{L}^2(\Omega)} \frac{|(\mathbf{u} - \mathbf{u}_h, \mathbf{w})_{\mathbf{L}^2(\Omega)}|}{\|\mathbf{w}\|_{\mathbf{L}^2(\Omega)}}. \tag{B.2.8}$$

Suppose that Ω is convex. We obtain (7.5.46) using the Aubin-Nitsche trick. Let $\mathbf{w} \in \mathbf{L}^2(\Omega)$. According to Proposition 7.2.1, it exists $(\mathbf{v}, q) \in \left((\mathbf{V} \cap H^2(\Omega)) \times (L^2_{zmv}(\Omega) \times H^1(\Omega)) \right)$ such that: $-\nu\Delta\mathbf{v} + \nabla q = \mathbf{w}$. By

integration by parts, letting $\delta\mathbf{u}_h = \mathbf{u} - \mathbf{u}_h$, we have for all $\mathbf{w} \in \mathbf{L}^2(\Omega)$:

$$\begin{aligned}(\mathbf{u}-\mathbf{u}_h,\mathbf{w})_{\mathbf{L}^2(\Omega)} &= \nu(\mathbf{u}-\mathbf{u}_h,\mathbf{v})_h - \nu\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}(\boldsymbol{\nabla}\mathbf{v}:\mathbf{n}_{f,\ell},\mathbf{u}-\mathbf{u}_h)_{\mathbf{L}^2(F_f)}\\ &\quad+\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}(q,(\mathbf{u}-\mathbf{u}_h)\cdot\mathbf{n}_{f,\ell})_{L^2(F_f)}.\end{aligned} \tag{B.2.9}$$

Summing (B.2.6) and (B.2.9), letting $\delta\mathbf{u}_h := \mathbf{u} - \mathbf{u}_h$, we deduce that for all $\mathbf{w} \in \mathbf{L}^2(\Omega)$ and for all $\mathbf{w}_h \in \mathbf{V}_{cr}$:

$$\begin{aligned}(\mathbf{u}-\mathbf{u}_h,\mathbf{w})_{\mathbf{L}^2(\Omega)} &= \nu(\mathbf{u}-\mathbf{u}_h,\mathbf{v}-\mathbf{w}_h)_h\\ &\quad+\nu\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}\Big((\boldsymbol{\nabla}\mathbf{u}:\mathbf{n}_{f,\ell},\mathbf{w}_h)_{\mathbf{L}^2(F_f)}-(\boldsymbol{\nabla}\mathbf{v}:\mathbf{n}_{f,\ell},\delta\mathbf{u}_h)_{\mathbf{L}^2(F_f)}\Big)\\ &\quad+\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}\Big((q,\delta\mathbf{u}_h\cdot\mathbf{n}_{f,\ell})_{L^2(F_f)}+(p,\mathbf{w}_h\cdot\mathbf{n}_{f,\ell})_{L^2(F_f)}\Big).\end{aligned} \tag{B.2.10}$$

We now consider $\mathbf{w}_h = \Pi_h\mathbf{v}$. By Cauchy-Schwarz, and using (7.5.45) and (7.5.38), we have:

$$|\nu(\mathbf{u}-\mathbf{u}_h,\mathbf{v}-\Pi_h\mathbf{v})_h| \le \nu\|\mathbf{u}-\mathbf{u}_h\|_h\,\|\mathbf{v}-\Pi_h\mathbf{v}\|_h \lesssim \nu\,\sigma^{4d}\,h^2\,\|(\mathbf{u},p)\|_\nu\,|\mathbf{v}|_{\mathbf{H}^2(\Omega)}.$$

We deduce, using Proposition 7.2.1 that:

$$|\nu(\mathbf{u}-\mathbf{u}_h,\mathbf{v}-\Pi_h\mathbf{v})_h| \lesssim \sigma^{4d}\,h^2\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{w}\|_{\mathbf{L}^2(\Omega)}. \tag{B.2.11}$$

We apply (7.4.2) to all the components of $\boldsymbol{\nabla}\mathbf{u} \in \mathbb{H}^1(\Omega)$ and $\mathbf{w}_h \in \mathbf{X}_{0,h}$:

$$\begin{aligned}\nu\left|\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}(\boldsymbol{\nabla}\mathbf{u}:\mathbf{n}_{f,\ell},\Pi_h\mathbf{v})_{\mathbf{L}^2(F_f)}\right| &= \nu\left|\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}(\boldsymbol{\nabla}\mathbf{u}:\mathbf{n}_{f,\ell},\mathbf{v}-\Pi_h\mathbf{v})_{\mathbf{L}^2(F_f)}\right|,\\ &\le \nu\,\sigma\,h\,|\mathbf{u}|_{\mathbf{H}^2(\Omega)}\,\|\mathbf{v}-\Pi_h\mathbf{v}\|_h.\end{aligned}$$

Using (7.5.38) and Proposition 7.2.1, it holds:

$$\nu\left|\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}(\boldsymbol{\nabla}\mathbf{u}:\mathbf{n}_{f,\ell},\Pi_h\mathbf{v})_{\mathbf{L}^2(F_f)}\right| \lesssim \nu\sigma^{2d}\,h^2\,|\mathbf{u}|_{\mathbf{H}^2(\Omega)}\,|\mathbf{v}|_{\mathbf{H}^2(\Omega)} \lesssim \sigma^{2d}\,h^2\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{w}\|_{\mathbf{L}^2(\Omega)}. \tag{B.2.12}$$

We apply (7.4.2) to all the components of $\boldsymbol{\nabla}\mathbf{v} \in \mathbb{H}^1(\Omega)$ and $\delta\mathbf{u}_h \in \mathbf{H}_0^1(\Omega)+\mathbf{X}_{0,h}$. Using (7.5.45) and Proposition 7.2.1, we obtain:

$$\nu\left|\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}(\boldsymbol{\nabla}\mathbf{v}:\mathbf{n}_{f,\ell},\delta\mathbf{u}_h)_{\mathbf{L}^2(F_f)}\right| \lesssim \sigma^{2d}\,h^2\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{w}\|_{\mathbf{L}^2(\Omega)}, \tag{B.2.13}$$

Summing (B.2.12) and (B.2.13), we get:

$$\nu\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}\Big((\boldsymbol{\nabla}\mathbf{u}:\mathbf{n}_{f,\ell},\mathbf{w}_h)_{\mathbf{L}^2(F_f)}-(\boldsymbol{\nabla}\mathbf{v}:\mathbf{n}_{f,\ell},\delta\mathbf{u}_h)_{\mathbf{L}^2(F_f)}\Big) \lesssim \sigma^{2d}\,h^2\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{w}\|_{\mathbf{L}^2(\Omega)}. \tag{B.2.14}$$

Samely, one proves that the two last terms in the right-hand side of (B.2.10) are bounded like:

$$\sum_{\ell\in\mathcal{I}_K}\sum_{f\in\mathcal{I}_{F,\ell}}\Big((q,\delta\mathbf{u}_h\cdot\mathbf{n}_{f,\ell})_{L^2(F_f)}+(p,\mathbf{w}_h\cdot\mathbf{n}_{f,\ell})_{L^2(F_f)}\Big) \lesssim \sigma^{2d}\,h^2\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{w}\|_{\mathbf{L}^2(\Omega)}. \tag{B.2.15}$$

Summing the contributions (B.2.11), (B.2.14) and (B.2.15), we deduce from (B.2.10) that:

$$|(\mathbf{u}-\mathbf{u}_h,\mathbf{w})_{\mathbf{L}^2(\Omega)}| \lesssim \sigma^{4d}\,h^2\,\|(\mathbf{u},p)\|_\nu\,\|\mathbf{w}\|_{\mathbf{L}^2(\Omega)}.$$

Applying this result to (B.2.8), we get (7.5.46).

B.3 Symmetric MPFA and nonconforming finite elements

When we develop the condition (7.6.2), we obtain:

$$\frac{1}{|K_{i-1}|}\mathcal{S}_{i-1,i-1}\cdot\mathcal{S}_{i-1,i}\,\tilde{q}_{i-1}+\left(\frac{1}{|K_{i-1}|}+\frac{1}{|K_i|}\right)|\mathcal{S}_{i,i}|^2\,\tilde{q}_i+\frac{1}{|K_i|}\mathcal{S}_{i,i}\cdot\mathcal{S}_{i,i+1}\,\tilde{q}_{i+1}$$

$$=-\frac{1}{|K_i|}\mathcal{S}_{i,i}\cdot\mathcal{S}_{i,S}\,\overline{q}_i-\frac{1}{|K_{i-1}|}\mathcal{S}_{i-1,i}\cdot\mathcal{S}_{0,i-1}\,\overline{q}_{i-1}.$$

Let $\widetilde{Q}=(\tilde{q}_i)_{i\in\mathcal{I}_S}$ and $\overline{Q}=(\overline{q}_i)_{i\in\mathcal{I}_S}$. We call $\mathbb{K}\in\mathbb{R}^{N_{S,0}\times N_{S,0}}$ and $\mathbb{K}_0\in\mathbb{R}^{N_{S,0}\times N_{S,0}}$ the matrices such that for all $i\in\mathcal{I}_S$:

$$\mathbb{K}_{i,i}=\left(\frac{1}{|K_{i-1}|}+\frac{1}{|K_i|}\right)|\mathcal{S}_{i,i}|^2,\quad \mathbb{K}_{i,i-1}=\frac{1}{|K_{i-1}|}\mathcal{S}_{i-1,i-1}\cdot\mathcal{S}_{i-1,i},\quad \mathbb{K}_{i,i+1}=\mathbb{K}_{i+1,i},$$

$$(\mathbb{K}_0)_{i,i}=\frac{1}{|K_i|}\mathcal{S}_{i,i}\cdot\mathcal{S}_{i,S},\quad (\mathbb{K}_0)_{i,i-1}=\frac{1}{|K_{i-1}|}\mathcal{S}_{i-1,i}\cdot\mathcal{S}_{i-1,S}.$$

The equations (7.6.2) are then equivalent to solving the following linear system:

$$\mathbb{K}\widetilde{P}=-\mathbb{K}_0\overline{P} \tag{B.3.1}$$

Let $\widetilde{\mathcal{M}}_S$ be the polygon formed by the midpoints $(M_i)_{i\in\mathcal{I}_{S,0}}$ and faces $(\widetilde{F}_{i,S_0})_{i\in\mathcal{I}_{S,0}}$ such that $\widetilde{F}_{i,S_0}=M_iM_{i+1}$.

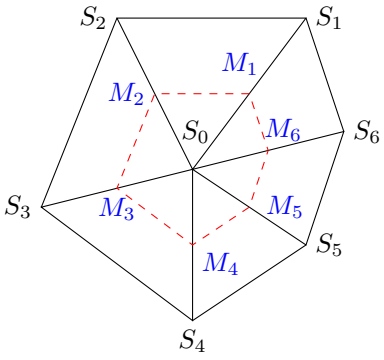


Figure B.4: Macro-element $\widetilde{\mathcal{M}}_0$.

The linear system (B.3.1) corresponds to the linear system obtained the P^1_{nc} finite elements of the following problem:

$$\begin{cases}\text{Find } p\in H^1(\widetilde{\mathcal{M}}_0) \text{ such that:}\\ -\Delta p &=& 0 & \text{in } \widetilde{\mathcal{M}}_0,\\ p_{|\widetilde{F}_{i,S_0}} &=& \overline{p}_i & \text{for all } i\in\mathcal{I}_{S,0}.\end{cases}$$

There exists a unique solution. By construction, according to 7.6.1, if p is linear, then the auxiliary pressures $(\tilde{p}_i)_{i\in\mathcal{I}_{S,0}}$ correspond to the values of p located at the third of the edges $(F_i)_{i\in\mathcal{I}_{S,0}}$.

B.4 Nonhomogeneous Dirichlet boundary conditions

We now solve the classical incompressible Stokes model, with non zero Dirichlet boundary conditions. Hence, we focus on Problem $(7.2.1)_{NH}$ with $g=0$. The variational formulation reads:

$$\begin{cases}\text{Find } (\mathbf{u},p)\in\mathbf{H}^1(\Omega)\times L^2_{zmv}(\Omega) \text{ such that}\\ \nu(\mathbf{u},\mathbf{v})_{\mathbf{H}^1_0(\Omega)}-(p,\operatorname{div}\mathbf{v})_{L^2(\Omega)} &=& \langle\mathbf{f},\mathbf{v}\rangle_{\mathbf{H}^{-1}(\Omega),\mathbf{H}^1_0(\Omega)} & \forall\mathbf{v}\in\mathbf{H}^1_0(\Omega),\\ (q,\operatorname{div}\mathbf{u})_{L^2(\Omega)} &=& 0 & \forall q\in L^2_{zmv}(\Omega),\\ \mathbf{u} &=& \mathbf{g} \text{ on } \partial\Omega,\end{cases} \tag{B.4.1}$$

where $\mathbf{g} \in \mathbf{H}^{\frac{1}{2}}(\partial\Omega)$ is some boundary data such that $\mathbf{g}\cdot\mathbf{n} \in L^2_{zmv}(\partial\Omega)$. Let $\mathbf{u_g} \in \mathbf{H}^1(\Omega)$ be such that $\mathbf{u_g} = \mathbf{g}$ on $\partial\Omega$ and $-\Delta\mathbf{u_g} = 0$ in Ω. Then consider $(\mathbf{u}_0, p) \in \mathbf{H}^1_0(\Omega) \times L^2_{zmv}(\Omega)$ the unique solution to $(7.2.1)_H$ with data $(\mathbf{f} + \nu\Delta\mathbf{u_g}, -\operatorname{div}\mathbf{u_g})$:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{u}_0, p) \in \mathcal{X} \text{ such that for all } (\mathbf{v}, q) \in \mathcal{X} \\ \quad a((\mathbf{u}_0, p), (\mathbf{v}, q)) = \langle \mathbf{f}, \mathbf{v}\rangle_{\mathbf{H}^{-1}(\Omega),\mathbf{H}^1_0(\Omega)} + (q, \operatorname{div}\mathbf{u_g})_{L^2(\Omega)}, \end{array}\right. \tag{B.4.2}$$

where we use that $(\mathbf{u_g}, \mathbf{v})_{\mathbf{H}^1_0(\Omega)} = 0$. Defining $\mathbf{u} = \mathbf{u_g} + \mathbf{u}_0 \in \mathbf{H}^1(\Omega)$, we find that $(\mathbf{u}, p)$ is solution to Problem (B.4.1). Uniqueness and continuous dependence with respect to the data are easily obtained. The next step is to replace $(\mathbf{v}, q)$ by $T((\mathbf{v}, q)) = (\lambda\mathbf{v} - \nu^{-1}\mathbf{v}_q, -\lambda q)$. One finds by orthogonality that for all $(\mathbf{v}, q) \in \mathcal{X}$:

$$\begin{array}{c} \nu\lambda(\mathbf{u}_0, \mathbf{v})_{\mathbf{H}^1_0(\Omega)} - \lambda(p, \operatorname{div}\mathbf{v})_{L^2(\Omega)} + \nu^{-1}(p, q)_{L^2(\Omega)} + \lambda(q, \operatorname{div}\mathbf{u}_0)_{L^2(\Omega)} \\ = \lambda\langle \mathbf{f}, \mathbf{v}\rangle_{\mathbf{H}^{-1}(\Omega),\mathbf{H}^1_0(\Omega)} - \nu^{-1}\langle \mathbf{f}, \mathbf{v}_q\rangle_{\mathbf{H}^{-1}(\Omega),\mathbf{H}^1_0(\Omega)} - \lambda(q, \operatorname{div}\mathbf{u_g})_{L^2(\Omega)}. \end{array}$$

Hence, a new variational formulation for nonhomogeneous boundary Dirichlet conditions is:
Find $(\mathbf{u}, p) \in \mathbf{H}^1(\Omega) \times L^2_{zmv}(\Omega)$ such that for all $(\mathbf{v}, q) \in \mathcal{X}$

$$\begin{array}{c} \nu\lambda(\mathbf{u}, \mathbf{v})_{\mathbf{H}^1_0(\Omega)} - \lambda(p, \operatorname{div}\mathbf{v})_{L^2(\Omega)} + \nu^{-1}(p, q)_{L^2(\Omega)} + \lambda(q, \operatorname{div}\mathbf{u})_{L^2(\Omega)} \\ = \lambda\langle \mathbf{f}, \mathbf{v}\rangle_{\mathbf{H}^{-1}(\Omega),\mathbf{H}^1_0(\Omega)} + \nu^{-1}(z_{\mathbf{f}}, q)_{L^2(\Omega)}. \end{array} \tag{B.4.3}$$

This is the same variational formulation as the one with homogeneous Dirichlet boundary conditions, except that $\mathbf{u} \in \mathbf{H}^1(\Omega)$ and $\mathbf{u}_{|\partial\Omega} = \mathbf{g}$.

Regarding the numerical algorithms, if $z_{\mathbf{f}}$ is known, one solves a linear system like

$$\left\{\begin{array}{l} \text{Find } (\underline{U}, \underline{P}) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p} \text{ such that:} \\ \begin{array}{rcl} \nu\lambda\mathbb{A}\,\underline{U} - \lambda\mathbb{B}^T\underline{P} & = & \lambda\underline{F} - \nu\lambda\mathbb{A}\,\underline{U}_{NH} \\ \lambda\mathbb{B}\,\underline{U} + \nu^{-1}\mathbb{M}\,\underline{P} & = & \nu^{-1}\mathbb{M}\,\underline{Z} - \lambda\mathbb{B}\,\underline{U}_{NH} \end{array} \end{array}\right. \tag{B.4.4}$$

where, in the right-hand side, $\underline{U}_{NH} \in \mathbb{R}^{N_u}$ accounts for $\mathbf{u_g}$. This is completely similar to (7.7.11). While, if $z_{\mathbf{f}}$ is not known, starting from an initial guess $\underline{P}^{-1} \in \mathbb{R}^{N_p}$, for $n = 0, 1, \cdots$, one solves linear systems like

$$\left\{\begin{array}{l} \text{Find } (\underline{U}^n, \underline{P}^n) \in \mathbb{R}^{N_u} \times \mathbb{R}^{N_p} \text{ such that:} \\ \begin{array}{rcl} \nu\lambda\mathbb{A}\,\underline{U}^n - \lambda\mathbb{B}^T\underline{P}^n & = & \lambda\underline{F}_u - \nu\lambda\mathbb{A}\,\underline{U}_{NH} \\ \lambda\mathbb{B}\,\underline{U}^n + \nu^{-1}\mathbb{M}\,\underline{P}^n & = & \nu^{-1}\mathbb{M}\,\underline{P}^{n-1} - \lambda\mathbb{B}\,\underline{U}_{NH} \end{array} \end{array}\right. . \tag{B.4.5}$$

Interestingly, one recovers the same results as those of Theorem 7.7.2, because one finds identical iterating matrices, cf. (7.7.14) and (7.7.15).

The (conforming) discretization of Problem (B.4.3) with $\mathbf{P}^k \times P^{k-1}_{disc}$ finite element reads:

$$\left\{\begin{array}{l} \text{Find } (\mathbf{u}_h, p_h) \in \mathbf{V}^k_h \times Q^{k-1}_h \text{ s.t. for all } (\mathbf{v}_h, q_h) \in \mathbf{V}^k_{0,h} \times Q^{k-1}_h \\ \nu\,\lambda(\mathbf{u}_h, \mathbf{v}_h)_{\mathbf{H}^1_0(\Omega)} - \lambda(p_h, \operatorname{div}\mathbf{v}_h)_{L^2(\Omega)} = \lambda\langle \mathbf{f}, \mathbf{v}_h\rangle_{\mathbf{H}^{-1}(\Omega),\mathbf{H}^1_0(\Omega)}, \\ \lambda(q_h, \operatorname{div}\mathbf{u}_h)_{L^2(\Omega)} + \nu^{-1}\,(p_h, q_h)_{L^2(\Omega)} = \nu^{-1}\,(z_{\mathbf{f}}, q_h)_{L^2(\Omega)}, \\ \mathbf{u}_{h|\partial\Omega} = (\Pi_{h,cg}\mathbf{u_g})_{|\partial\Omega}. \end{array}\right. \tag{B.4.6}$$

For nonhomogeneous boundary Dirichlet conditions, one can study the error estimates by introducing $\mathbf{u}_{0,h} = \mathbf{u}_h - \Pi_{h,cg}\mathbf{u_g} \in \mathbf{V}^k_{0,h}$ and using the results that have been obtained for the homogeneous boundary conditions. First, in the estimate (7.7.23) of Proposition 7.7.5, we change the term $\|z_{\mathbf{f}} - \tilde{z}_{\mathbf{f},h}\|_{L^2(\Omega)}$ in the right-hand side of the two equations into $\|z_{\mathbf{f}} - \tilde{z}_{\mathbf{f},h}\|_{L^2(\Omega)} + \lambda\nu\|\operatorname{div}(\mathbf{u_g} - \Pi_{h,cg}\mathbf{u_g})\|_{L^2(\Omega)}$. Second, in the estimate (7.7.25) of Theorem 7.7.3, we add $\sqrt{\lambda}\|\operatorname{div}(\mathbf{u_g} - \Pi_{h,cg}\mathbf{u_g})\|_{L^2(\Omega)}$ to the right-hand side.

Bibliography

Articles published in peer-reviewed journals

[A1] A. Buffa, P. Ciarlet Jr., and E. Jamelot. Solving electromagnetic eigenvalue problems in polyhedral domains with nodal finite elements. *Numer. Math.*, 113(4):497–518, 2009.

[A2] P. Ciarlet Jr., L. Giret, E. Jamelot, and F. D. Kpadonou. Numerical analysis of the mixed finite element method for the neutron diffusion eigenproblem with heterogeneous coefficients. *ESAIM: M2AN*, 52(5):2003–2035, 2018.

[A3] P. Ciarlet Jr. and E. Jamelot. Continuous Galerkin methods for solving the time-dependent Maxwell equations in 3D geometries. *Journal of Computational Physics*, 226(1):1122–1135, 2007.

[A4] P. Ciarlet Jr. and E. Jamelot. Variational methods for solving numerically magnetostatic systems. *Advances in Computational Mathematics*, 50(5), 2024.

[A5] P. Ciarlet Jr. and E. Jamelot. Explicit T-coercivity for the Stokes problem: a coercive finite element discretization. *Computers & Mathematics with Applications*, 188:137–159, 2025.

[A6] P. Ciarlet Jr., E. Jamelot, and F. Kpadonou. Domain decomposition methods for the diffusion equation with low-regularity solution. *Computers & Mathematics with Applications*, 74(10):2369–2384, 2017.

[A7] S. Dellacherie, E. Jamelot, and O. Lafitte. A simple monodimensional model coupling an enthalpy production equation and a neutronic diffusion equation. *Applied Mathematics Letters*, 62:35–41, 2016.

[A8] E. Jamelot. Éléments finis nodaux pour les équations de Maxwell. *C. R. Acad. Sci. Paris, Ser. I*, 339(11):809–814, 2004. Corrected in *C. R. Acad. Sci. Paris, Ser. I*, 340(5):409–410, 2005.

[A9] E. Jamelot. Stability estimates for solving Stokes problem with nonconforming finite elements. *Comptes Rendus. Mathématique*, 363:115–137, 2025.

[A10] E. Jamelot, A.-M. Baudron, and J.-J. Lautard. Domain Decomposition for the SPN Solver MINOS. *Transport Theory and Statistical Physics*, 41:495–512, 2012.

[A11] E. Jamelot and P. Ciarlet Jr. Fast non-overlapping Schwarz domain decomposition methods for solving the neutron diffusion equation. *Journal of Computational Physics*, 241:445–463, 2013.

[A12] E. Jamelot and F. Madiot. Numerical analysis of the neutron multigroup SPN equations. *Comptes Rendus. Mathématique*, 359(5):533–545, 2021.

[A13] M. A. Puscas, P.-E. Angeli, N. Nouaime, and E. Jamelot. Description and convergence order analysis of the Finite Element-Volume spatial discretization method. *International Journal for Numerical Methods in Fluids*, 2025.

Articles submitted or soon to be finalised

[W1] P.-E. Angeli, H. Bertrand, and E. Jamelot. Convergence analysis of the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ discretization in the TrioCFD code. Submitted to Computers & Mathematics With Applications, 2025.

[W2] E. Jamelot, M. Mroueh, and P. Omnes. A Priori Error Estimates for Discontinous Galerkin Finite Element Discretization of Stokes Problems with Low Regularity. Article in progress, 2025.

Peer-reviewed proceedings

[P1] P. Ciarlet Jr. and E. Jamelot. Continuous Galerkin methods for solving Maxwell equations in 3D geometries. In A. Bermúdez de Castro and D. Gómez and P. Quintela and P. Salgado, editor, *Numerical Mathematics and Advanced Applications, ENUMATH 2005*. Springer, 2005.

[P2] S. Dellacherie, E. Jamelot, O. Lafitte, and R. Mouhamad. Numerical results for the coupling of a simple neutronics diffusion model and a simple hydrodynamics low mach number model without coupling codes. In *SYNASC 2016*. West University of Timisoara, 2016.

[P3] É. Chénier, E. Jamelot, C. Le Potier, and A. Peitavy. Improved Crouzeix-Raviart Scheme for the Stokes Problem. In E. Franck and J. Fuhrmann and V. Michel-Dansac and L. Navoret, editor, *Finite Volumes for Complex Applications X—Volume 1, Elliptic and Parabolic Problems*, pages 245–253, Cham, 2023. Springer Nature Switzerland.

[P4] É. Chénier, E. Jamelot, C. Le Potier, and A. Peitavy. Improved Crouzeix-Raviart scheme for the Stokes and Navier-Stokes problem. *ESAIM: ProcS*, 76:20–34, 2024.

[P5] L. Giret, P. Ciarlet Jr., and E. Jamelot. Critical Computation with Finite Element Method on Non-Conforming Meshes. In *M&C 2017*. Korea Nuclear Society, 2017.

[P6] E. Jamelot, P. Ciarlet Jr., A.-M. Baudron, and J.-J. Lautard. Domain decomposition for the neutron SPN equations. In J. Erhel, M.J. Gander, L. Halpern, G. Pichot, T. Sassi and O. Widlund, editor, *Domain Decomposition Methods in Science and Engineering XXI*, volume 98 of *Lecture Notes in Computational Science and Engineering*. Springer, 2014.

[P7] E. Jamelot, P. Ciarlet Jr., and S. Sauter. Stability of the $\mathbf{P}^1_{nc} \times (P^0 + P^1)$ element. In A. Sequeira, A. Silvestre, S. S. Valtchev, and J. Janela, editors, *Numerical Mathematics and Advanced Applications ENUMATH 2023, Volume 1*, pages 494–503, Cham, 2025. Springer Nature Switzerland.

[P8] E. Jamelot, J. Dubois, J.-J. Lautard, C. Calvin, and A.-M. Baudron. High performance 3D neutron transport on petascale and hybrid architectures within APOLLO3 code. In *M&C 2011*. International Atomic Energy Agency, 2011.

[P9] M. Mroueh, E. Jamelot, and P. Omnes. Discontinuous Galerkin schemes for the Stokes problem. In *SNA + MC 2024*, volume 302, Paris, France, Oct 2024. EPJ Web of Conferences.

[P10] D. Schneider, F. Dolci, F. Gabriel, J.-M. Palau, M. Guillo, and B. P. et al. APOLLO3: CEA/DEN deterministic multi-purpose code for reactor physics analysis. In *PHYSOR 2016 – Unifying Theory and Experiments in the 21st Century*, 2016.

Lecture notes and PhD

[B1] P. Ciarlet Jr. and E. Jamelot. Outils mathématiques et algorithmiques pour le calcul scientifique. ENSTA, `https://hal.science/hal-04311490v1`, 2019.

[B2] E. Jamelot. *Continuous Galerkin methods for solving Maxwell's equations*. PhD thesis, École Polytechnique, Palaiseau, France, 2005.

GitHub digital models

[G1] E. Jamelot. `https://github.com/cea-trust-platform/Stokes_NCFEM`, 2022.

[G2] E. Jamelot. `https://github.com/cea-trust-platform/Stokes_ExplicitTC`, 2024.

References

[1] Y. Achdou, C. Japhet, Y. Maday, and F. Nataf. A new cement to glue non-conforming grids with Robin interface conditions: The finite volume case. *Numer. Math.*, 92(4):593–620, 2002.

[2] M. Ainsworth, A. Allendes, G. Barrenechea, and R. Rankin. Computable error bounds for nonconforming Fortin–Soulie finite element approximation of the Stokes problem. *IMA Journal of Numerical Analysis*, 32(2):417–447, 2011.

[3] M. Ainsworth, A. Allendes, G. Barrenechea, and R. Rankin. On the Adaptive Selection of the Parameter in Stabilized Finite Element Approximations. *SIAM Journal on Numerical Analysis*, 51(3):1585–1609, 2013.

[4] P. Alfeld. A trivariate Clough-Tocher scheme for tetrahedral data. *Comput. Aided Geom. Design*, 1:169–181, 1984.

[5] A. Allendes, G. Barrenechea, and C. Naranjo. A divergence-free low-order stabilized finite element method for a generalized steady state Boussinesq problem. *Computer Methods in Applied Mechanics and Engineering*, 340:90–120, 2018.

[6] A. Alonso and A. Russo. Spectral approximation of variationally-posed eigenvalue problems by nonconforming methods. *Journal of Computational and Applied Mathematics*, 223(1):177–197, 2009.

[7] C. Amrouche, C. Bernardi, M. Dauge, and V. Girault. Vector potentials in three-dimensional non-smooth domains. *Math. Meth. Appl. Sci.*, 21:823–864, 1998.

[8] C. Amrouche, P. Ciarlet, and C. Mardare. On a lemma of Jacques-Louis Lions and its relation to other fundamental results. *Journal de Mathématiques Pures et Appliquées*, 104(2):207–226, 2015.

[9] P.-E. Angeli. Overview of the TrioCFD Code: Main Features, V&V Procedures and Typical Applications to Nuclear Engineering. In *16th International Topical Meeting on Nuclear Reactor Thermalhydraulics (NURETH-16)*, Chicago, USA, 2015.

[10] P.-E. Angeli, U. Bieder, and G. Fauchet. Overview of the TrioCFD code: main features, V&V procedures and typical applications to nuclear engineering. In *Proceedings of 16th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-16)*, 2015.

[11] P.-E. Angeli, M. Puscas, G. Fauchet, and A. Cartalade. FVCA8 benchmark for the Stokes and Navier-Stokes equations with the TrioCFD code. In C. Cancès and P. Omnes, editor, *Finite Volumes for Complex Applications VIII – Methods and Theoretical Aspects*, volume 199 of *Proceedings in Mathematics & Statistics*, pages 181–302. Springer, 2017.

[12] T. Apel, V. Kempf, A. Linke, and C. Merdon. A nonconforming pressure-robust finite element method for the Stokes equations on anisotropic meshes. *IMA Journal of Numerical Analysis*, 42(1):392–416, 2022.

[13] T. Apel, S. Nicaise, and J. Schöberl. Crouzeix-Raviart type finite elements on anisotropic meshes. *Numer. Math.*, 89(2):193–223, 2001.

[14] J.-P. Argaud, B. Bouriquet, F. D. Caso, H. Gong, Y. Maday, and O. Mula. Sensor placement in nuclear reactors based on the generalized empirical interpolation method. *Journal of Computational Physics*, 363:354–370, 2018.

[15] D. Arnold, F. Brezzi, and M. Fortin. A stable finite element for the Stokes equations. *Calcolo*, 21:337–344, 1984.

[16] F. Assous, P. Ciarlet Jr., and S. Labrunie. *Mathematical Foundations of Computational Electromagnetism*, volume 198 of *Applied Mathematical Sciences*. Springer, 2018.

[17] F. Assous, P. Degond, E. Heintzé, P.-A. Raviart, and J. Segré. On a finite element method for solving the three-dimensional Maxwell equations. *J. Comput. Phys.*, 109:222–237, 1993.

[18] I. Babuška and J. Osborn. Eigenvalue problems. In *Finite Element Methods (Part 1)*, volume 2 of *Handbook of Numerical Analysis*, pages 641–787. Elsevier, 1991.

[19] S. Badia and R. Codina. A combined nodal continuous-discontinuous finite element formulation for the Maxwell problem. *Applied Mathematics and Computation*, 218:4276–4294, 2011.

[20] S. Badia and R. Codina. A nodal-based finite element approximation of the Maxwell problem suitable for singular solutions. *SIAM Journal on Numerical Analysis*, 50:398–417, 2012.

[21] À. Baran and G. Stoyan. Gauss-Legendre elements: a stable, higher order non-conforming finite element family. *Computing*, 79(1):1–21, 2007.

[22] M. Barrault, B. Lathuilière, P. Ramet, and J. Roman. Efficient parallel resolution of the simplified transport equations in mixed-dual formulation. *Journal of Computational Physics*, 230(5):2004–2020, 2011.

[23] G. Barrenechea, E. Burman, E. Cáceres, and J. Guzmán. Continuous interior penalty stabilization for divergence-free finite element methods. *IMA Journal of Numerical Analysis*, 2023.

[24] G. Barrenechea and F. Valentin. An unusual stabilized finite element method for a generalized Stokes problem. *Numer. Math.*, 92:653–677, 2002.

[25] G. Barrenechea and F. Valentin. Consistent Local Projection Stabilized Finite Element Methods. *SIAM Journal on Numerical Analysis*, 48(5):1801–1825, 2010.

[26] M. Barré and P. Ciarlet Jr. The T-coercivity approach for mixed problems. *Comptes rendus. Mathématique*, 362:1051–1088, 2024.

[27] R. Barthelmé, P. Ciarlet Jr., and E. Sönnendrücker. Generalized formulations of Maxwell's equations for numerical Vlasov-Maxwell simulations. *Mathematical Models and Methods in Applied Sciences*, 17(5):657–680, 2007.

[28] A.-M. Baudron and J.-J. Lautard. MINOS: A Simplified P_N solver for core calculations. *Nuclear Science and Engineering*, 155:250–263, 2007.

[29] A.-M. Baudron and J.-J. Lautard. SP_N core calculations in the APOLLO3 System. In *Mathematics and Computational Methods Applied to Nuclear Science and Engineering (M&C 2011)*. Latin American Section (LAS) / American Nuclear Society (ANS), 2011.

[30] A.-M. Baudron, J.-J. Lautard, Y. Maday, and O. Mula. MINARET: Towards a time-dependent neutron transport parallel solver. In Array, editor, *SNA + MC 2013 - Joint International Conference on Supercomputing in Nuclear Applications + Monte Carlo*, page 04103, 2014.

[31] A.-M. Baudron, J.-J. Lautard, Y. Maday, and O. Mula. The parareal in time algorithm applied to the kinetic neutron diffusion equation. In J. Erhel, M.J. Gander, L. Halpern, G. Pichot, T. Sassi and O. Widlund, editor, *Domain Decomposition Methods in Science and Engineering XXI*, volume 98 of *Lecture Notes in Computational Science and Engineering*, pages 437–445. Springer, 2014.

[32] F. Ben Belgacem and C. Bernardi. Spectral element discretization of the Maxwell equations. *Math. Comp.*, 68:1497–1520, 1999.

[33] F. Ben Belgacem and S. Brenner. Some nonstandard finite element estimates with applications to 3D Poisson and Signorini problems. *Electronic Transactions in Numerical Analysis*, 12:134–148, 2001.

[34] A. Bermúdez, P. Gamallo, M. Nogueiras, and R. Rodriguez. Approximation of a structural acoustic vibration problem by hexahedral finite elements. *IMA J. Numer. Anal.*, 26:391–421, 2006.

[35] A. Bermúdez, D. Gómez, and P. Salgado. *Mathematical models and numerical simulation in electromagnetism*. Springer, Cham, Switzerland, 2014.

[36] A. Bermúdez, B. López-Rodriguez, R. Rodriguez, and P. Salgado. Equivalence between two finite element methods for the eddy current problem. *C. R. Acad. Sci. Paris, Ser. I*, 348:769–774, 2010.

[37] A. Bermúdez, M. Piñiero, R. Rodriguez, and P. Salgado. Analysis of an ungauged $T, \phi - \phi$ formulation of the eddy current problem with currents and voltage excitations. *Math. Mod. Num. Anal.*, 51:2487–2509, 2017.

[38] A. Bermúdez, M. Piñiero, and P. Salgado. Mathematical and numerical analysis of a transient magnetic model with voltage drop excitations. *Computers Math. Applic.*, 76:2710–2722, 2018.

[39] A. Bermúdez, R. Rodriguez, and P. Salgado. Numerical treatment of realistic boundary conditions for the eddy current problem in an electrode via Lagrange multipliers. *Math. Comp.*, 74:123–151, 2005.

[40] A. Bermúdez, R. Rodriguez, and P. Salgado. A finite element method for the magnetostatic problem in terms of scalar potentials. *SIAM Journal on Numerical Analysis*, 46:1338–1363, 2008.

[41] C. Bernardi, M. Costabel, M. Dauge, and V. Girault. Continuity properties of the inf-sup constant for the divergence. *SIAM Journal on Mathematical Analysis*, 48(2):1250–1271, 2016.

[42] C. Bernardi and F. Hecht. More pressure in the finite element discretisation of the Stokes problem. *ESAIM: M2AN*, 34(5):953–980, 2000.

[43] C. Bernardi, F. Hecht, and O. Pironneau. Coupling Darcy with Stokes equations for porous media with cracks. *ESAIM: M2AN*, 39(1):7–35, 2005.

[44] C. Bernardi and G. Raugel. Méthodes d'éléments finis mixtes pour les équations de Stokes et de Navier-Stokes dans un polygone non convexe. *Calcolo*, 18:255–291, 1981.

[45] C. Birdsall and A. Langdon. *Plasmas Physics via Computer Simulation*. Mac Graw-Hill, New-York, 1985.

[46] M. Birman and M. Solomyak. Maxwell operator in regions with nonsmooth boundaries. *Sib. Math. J.*, 28:12–24, 1987.

[47] P. Bochev, C. Dohrmann, and M. D. Gunzburger. Stabilization of low-order mixed finite elements for the Stokes equations. *SIAM Journal on Numerical Analysis*, 44(1):82–20, 2006.

[48] D. Boffi. Compatible Discretizations for Eigenvalue Problems. In D. N. Arnold, P. B. Bochev, R. B. Lehoucq, R. A. Nicolaides, and M. Shashkov, editors, *Compatible Spatial Discretizations*, pages 121–142, New York, NY, 2006. Springer New York.

[49] D. Boffi. Finite element approximation of eigenvalue problems. *Acta Numerica*, 19:1–120, 2010.

[50] D. Boffi, F. Brezzi, and M. Fortin. *Mixed and hybrid finite element methods and applications*. Springer-Verlag, 2013.

[51] D. Boffi, F. Brezzi, and L. Gastaldi. On the convergence of eigenvalues for mixed formulations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci., Ser. 4*, 25(1-2):131–154, 1997.

[52] D. Boffi, F. Brezzi, and L. Gastaldi. On the problem of spurious eigenvalues in the approximation of linear elliptic problems in mixed form. *Math. Comp.*, 69(229):121–140, 2000.

[53] D. Boffi, D. Gallistl, F. Gardini, and L. Gastaldi. Optimal convergence of adaptive FEM for eigenvalue clusters in mixed form. *Math. Comp.*, 86:2213–2237, 2017.

[54] A. Bonito and J.-L. Guermond. Approximation of the eigenvalue problem for the time harmonic Maxwel system by continuous Lagrange finite elements. *Math. Comp.*, 80:1887–1910, 2011.

[55] A. Bonito, J.-L. Guermond, and F. Luddens. Regularity of the Maxwell equations in heterogeneous media and Lipschitz domains. *J. Math. Anal. Appl.*, 408:498–512, 2013.

[56] A.-S. Bonnet-Ben Dhia, C. Carvalho, and P. Ciarlet Jr. Mesh requirements for the finite element approximation of problems with sign-changing coefficients. *Numer. Math.*, 138:801–838, 2018.

[57] A.-S. Bonnet-Ben Dhia, L. Chesnel, and P. Ciarlet Jr. T-coercivity for scalar interface problems between dielectrics and metamaterials. *Math. Mod. Num. Anal.*, 46:1363–1387, 2012.

[58] A.-S. Bonnet-Ben Dhia, L. Chesnel, and P. Ciarlet Jr. T-coercivity for the Maxwell problem with sign-changing coefficients. *Communications in Partial Differential Equations*, 39:1007–1031, 2014.

[59] A.-S. Bonnet-Ben Dhia, L. Chesnel, and H. Haddar. On the use of T-coercivity to study the interior transmission eigenvalue problem. *C. R. Math. Acad. Sci. Paris*, 349:647–651, 2011.

[60] A.-S. Bonnet-Ben Dhia and P. Ciarlet Jr. Méthodes variationnelles pour l'analyse de problèmes non coercifs, 2023. M.Sc. AMS lecture notes (ENSTA-IPP).

[61] A.-S. Bonnet-Ben Dhia, P. Ciarlet Jr., and C. Zwölf. Time harmonic wave diffraction problems in materials with sign-shifting coefficients. *J. Comput. Appl. Math.*, 234:1912–1919, corrigendum p. 2616, 2010.

[62] A.-S. Bonnet-Ben Dhia, C. Hazard, and S. Lohrengel. A singular field method for the solution of Maxwell's equations in polyhedral domains. *SIAM Journal on Applied Mathematics*, 59:2028–2044, 1999.

[63] A. Bossavit. *Computational electromagnetism*. Academic Press, New York, USA, 1998.

[64] F. Boyer and P. Fabrie. *Mathematical tools for the study of the incompressible Navier-Stokes equations and related models*. Springer-Verlag, 2013.

[65] J. H. Bramble and J. E. Pasciak. A new approximation technique for div-curl systems. *Math. Comp.*, 73:1739–1762, 2003.

[66] C. Brennecke, A. Linke, C. Merdon, and J. Schöberl. Optimal and pressure independent L^2 velocity error estimates for a modified Crouzeix-Raviart Stokes element with BDM reconstructions. *Journal of Computational Mathematics*, 33(2):191–208, 2015.

[67] S. Brenner. A multigrid algorithm for the lowest-order Raviart-Thomas mixed trangular finite element method. *SIAM Journal on Numerical Analysis*, 29(3):647–678, 1992.

[68] S. Brenner. Poincaré-Friedrichs inequalities for piecewise H^1 functions. *SIAM Journal on Numerical Analysis*, 41(1):306–324, 2003.

[69] S. Brenner, F. Li, and L.-Y. Sung. A locally divergence-free interior penalty method for two dimensional curl–curl problems. *SIAM Journal on Numerical Analysis*, 46(3):1190–1211, 2008.

[70] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2011.

[71] F. Brezzi, J. Douglas, and L. D. Marini. Two families of mixed finite elements for second order elliptic problems. *Numer. Math.*, 47(2):217–235, 1985.

[72] A. Buffa. Remarks on the discretization of some noncoercive operator with applications to heterogeneous Maxwell equations. *SIAM Journal on Numerical Analysis*, 43:1–18, 2005.

[73] A. Buffa and S. H. Christiansen. The electric field integral equation on Lipschitz screens: definitions and numerical approximation. *Numer. Math.*, 94(2):229–267, 2003.

[74] A. Buffa and P. Ciarlet Jr. On traces for functional spaces related to Maxwell's equations. Part I: an integration by parts formula in Lipschitz polyhedra. *Math. Meth. Appl. Sci.*, 24:9–30, 2001.

[75] A. Buffa, M. Costabel, and M. Dauge. Anisotropic regularity results for Laplace and Maxwell operators in a polyhedron. *C. R. Acad. Sci. Paris, Série I*, 336:565–570, 2003.

[76] A. Buffa, M. Costabel, and C. Schwab. Boundary element methods for Maxwell's equations on non-smooth domains. *Numer. Math.*, 92(4):679–710, 2002.

[77] D.-Q. Bui, C. Japhet, Y. Maday, and P. Omnes. Coupling Parareal with optimized Schwarz waveform relaxation for parabolic problems. *SIAM Journal on Numerical Analysis*, 60(3):913–939, 2022.

[78] D.-Q. Bui, C. Japhet, and P. Omnes. Optimized Schwarz waveform relaxation method for the incompressible Stokes problem. *ESAIM: M2AN*, 58:1229–1261, 2024.

[79] E. Burman and P. Hansbo. Stabilized Crouzeix-Raviart element for the Darcy-Stokes problem. *Numerical Methods for Partial Differential Equations*, 21(5):986–997, 2005.

[80] J. Bussac and P. Reuss. *Traité de neutronique*. Hermann, 1985.

[81] C. Carstensen and S. Sauter. Critical functions and inf-sup stability of Crouzeix-Raviart elements. *Computers and Mathematics with Applications*, 108:12–23, 2022.

[82] C. Carstensen and S. Sauter. Crouzeix-Raviart triangular elements are inf-sup stable, 2022.

[83] N. Chaabane, V. Girault, B. Riviere, and T. Thompson. A stable enriched Galerkin element for the Stokes problem. *Applied Numerical Mathematics*, 132:1–21, 2018.

[84] X. Chen, M. J. Gander, and Y. Xu. Optimized Schwarz Methods with Elliptical Domain Decompositions. *J Sci Comput*, 86(27), 2006.

[85] L. Chesnel and P. Ciarlet Jr. T-coercivity and continuous Galerkin methods: application to transmission problems with sign changing coefficients. *Numer. Math.*, 124:1–29, 2013.

[86] D. Chicaud. *Analysis of time-harmonic electromagnetic problems in elliptic anisotropic media*. PhD thesis, Institut Polytechnique de Paris, 2021.

[87] D. Chicaud and P. Ciarlet Jr. Analysis of time-harmonic Maxwell impedance problems in anisotropic media. *SIAM Journal on Mathematical Analysis*, 55:1969–2000, 2023.

[88] D. Chicaud, P. Ciarlet Jr., and A. Modave. Analysis of variational formulations and low-regularity solutions for time-harmonic electromagnetic problems in complex anisotropic media. *SIAM Journal on Mathematical Analysis*, 53:2691–2717, 2021.

[89] A. J. Chorin. Numerical solution of the Navier-Stokes equations. *Mathematics of Computation*, 22(104):745–762, 1968.

[90] P. Ciarlet. Basic error estimates for elliptic problems. In *Finite Element Methods (Part 1)*, volume 2 of *Handbook of Numerical Analysis*, pages 17–351. Elsevier, 1991.

[91] P. Ciarlet Jr. Augmented formulations for solving Maxwell equations. *Comput. Methods Appl. Mech. Engrg.*, 194:559–586, 2005.

[92] P. Ciarlet Jr. T-coercivity: Application to the discretization of Helmhotz-like problems. *Computers & Mathematics with Applications*, 64(1):22–24, 2012.

[93] P. Ciarlet Jr. Analysis of the Scott-Zhang interpolation in the fractional order Sobolev spaces. *Journal of Numerical Mathematics*, 21(3):173–180, 2013.

[94] P. Ciarlet Jr. On the approximation of electromagnetic fields by edge finite elements. Part 1: sharp interpolation results for low-regularity fields. *Computers Math. Applic.*, 71:85–104, 2016.

[95] P. Ciarlet Jr. T-coercivity: a practical tool for the study of variational formulations in Hilbert spaces, 2025.

[96] P. Ciarlet Jr., M. H. Do, and F. Madiot. A posteriori error estimates for mixed finite element discretizations of the Neutron Diffusion equations. *ESAIM: M2AN*, 57(1):1–27, 2023.

[97] P. Ciarlet Jr. and É. Lunéville. *The finite element Method*. From theory to practice Numerical Methods in Engineering Series. WILEY, 2023.

[98] P. Ciarlet Jr., G. Garcia, and J. Zou. Résolution des équations de Maxwell dans des domaines prismatiques tridimensionnels. *C. R. Acad. Sci. Paris, Série I*, 339:721–726, 2004.

[99] P. Ciarlet Jr., C. Hazard, and S. Lohrengel. Les équations de Maxwell dans un polyèdre : un résultat de densité. *C. R. Acad. Sci. Paris, Ser. I*, 326:1305–1310, 1998.

[100] P. Ciarlet Jr. and G. Hechme. Computing electromagnetic eigenmodes with continuous Galerkin approximations. *Comput. Methods Appl. Mech. Eng.*, 198(2):358–365, 2008.

[101] P. Ciarlet Jr. and S. Labrunie. Numerical analysis of the generalized Maxwell equations (with an elliptic correction) for charged particle simulations. *Math. Models Meth. App. Sci.*, 19:1959–1994, 2009.

[102] P. Ciarlet Jr. and S. Labrunie. Numerical solution of Maxwell's equations in axisymmetric domains with the Fourier Singular Complement Method. *Differential Equations & Applications*, 3(1):113–155, 2011.

[103] P. Ciarlet Jr. and C. Scheid. Electrowetting of a 3D drop: numerical modelling with electrostatic vector fields. *Math. Mod. Num. Anal.*, 44:647–670, 2010.

[104] P. Ciarlet, Jr, H. Wu, and J. Zou. Edge element methods for Maxwell's equations with strong convergence for Gauss' laws. *SIAM Journal on Numerical Analysis*, 52:779–807, 2014.

[105] P. Ciarlet Jr. and J. Zou. Fully discrete finite element approaches for time-dependent Maxwell's equations. *Numer. Math.*, 82:193–219, 1999.

[106] X. Claeys. NonLocal Optimized Schwarz Method for the Helmholtz Equation with Physical Boundaries. *SIAM Journal on Mathematical Analysis*, 55(6):7490–7512, 2023.

[107] X. Claeys, F. Collino, P. Joly, and E. Parolin. A discrete domain decomposition method for acoustics with uniform exponential rate of convergence using non-local impedance operators. In *Domain Decomposition Methods in Science and Engineering XXV*, volume 138 of *LNCSE*, pages 310–317. Springer, 2020.

[108] X. Claeys, F. Collino, and B. Thierry. Integral equation based optimized Schwarz method for electromagnetics. In *Domain Decomposition Methods in Science and Engineering XXIV*, volume 125 of *LNCSE*, pages 187–194. Springer, 2018.

[109] R. W. Clough and J. L. Tocher. Finite element stiffness matrices for analysis of plates in bending. In *Conference on Matrix Methods in Structural Mechanics, Wright Patterson A.F.B, Dayton, USA*, pages 515–545, 1965.

[110] E. A. Coddington and N. N. Levinson. *Theory of Ordinary Differential Equations*. International Series in Pure and Applied Mathematics. McGraw-Hill, 1955.

[111] G. Cohen. *Higher-Order Numerical Methods for Transient Wave Equations*. Scientific Computation. Springer-Verlag, Berlin, 2002.

[112] M. Costabel. A coercive bilinear form for Maxwell's equations. *J. Math. An. Appl.*, 157:527–541, 1991.

[113] M. Costabel and M. Dauge. Un résultat de densité pour les équations de Maxwell régularisées dans un domaine lipschitzien. *C. R. Acad. Sci. Paris, Ser. I*, 327:849–854, 1998.

[114] M. Costabel and M. Dauge. Singularities of electromagnetic fields in polyhedral domains. *Arch. Ration. Mech. Anal.*, 151:221–276, 2000.

[115] M. Costabel and M. Dauge. Crack Singularities for General Elliptic Systems. *Mathematische Nachrichten*, 235(1):29–49, 2002.

[116] M. Costabel and M. Dauge. Weighted regularization of Maxwell equations in polyhedral domains. *Numer. Math.*, 93:239–277, 2002.

[117] M. Costabel and M. Dauge. *Computations of resonance frequencies for Maxwell equations in non-smooth domains*, pages 125–161. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[118] M. Costabel and M. Dauge. On the inequalities of Babuška-Aziz, Friedrichs and Horgan-Payne. *Arch. Rational Mech. Anal.*, 217:873–898, 2015.

[119] M. Costabel, M. Dauge, and S. Nicaise. Singularities of electromagnetic fields in polyhedral domains. *ESAIM: M2AN*, 33(3):627–649, 1999.

[120] E. Creusé, P. Dular, and S. Nicaise. About the gauge conditions arising in Finite Element magnetostatic problems. *Computers & Mathematics with Applications*, 77(6):1563–1582, 2019.

[121] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations I. *Revue française d'automatique, informatique, recherche opérationnelle. Mathématique*, 7(3):33–75, 1973.

[122] D.A. Di Pietro and J. Droniou. From Finite Elements to Hybrid High-Order methods. `https://arxiv.org/html/2503.00425v1`, 2025.

[123] E. Dari, R. Durán, and C. Padra. Error estimators for nonconforming finite element approximations of the Stokes problem. *Mathematics of Computation*, 64(211):1017–1033, 1995.

[124] M. Dauge. *Elliptic Boundary Value Problems on Corner Domains*. Springer-Verlag, 1988.

[125] M. Dauge. Benchmark computations for Maxwell equations for the approximation of highly singular solutions. `https://perso.univ-rennes1.fr/monique.dauge/core/index.html`, 2004.

[126] R. Dautray and J.-L. Lions. *Analyse Mathématique et Calcul Numérique pour les Sciences et les Techniques*. MASSON, 1987.

[127] S. Dellacherie. On a low Mach nuclear core model. *ESAIM: Proc.*, 35:79–106, 2012.

[128] S. Dellacherie and O. Lafitte. Une solution explicite monodimensionnelle d'un modèle simplifié de couplage stationnaire thermohydraulique-neutronique. *Annales mathématiques du Québec*, 41:221–264, 2016.

[129] A. Dello Russo and A. Alonso. Finite element approximation of Maxwell eigenproblems on curved Lipschitz polyhedral domains. *Appl. Numer. Math.*, 59:1796–1822, 2009.

[130] D. Di Pietro and A. Ern. *Mathematical Aspects of Discontinuous Galerkin Methods*. Springer, 2012.

[131] L. Diening, J. Storn, and T. Tscherpel. Fortin operator for Taylor-Hood element. *Numer. Math.*, 150(2):671–689, 2022.

[132] M. H. Do, P. Ciarlet Jr., and F. Madiot. Adaptive solution of the neutron diffusion equation with heteregenous coefficients using the mixed finite element method on structured meshes. *EPJ Web Conf.*, 247:02002, 2021.

[133] V. Dolean, M. J. Gander, and A. Kyriakis. Optimizing transmission conditions for multiple subdomains in the Magnetotelluric Approximation of Maxwell's equations. In *Domain Decomposition Methods in Science and Engineering XXVI*, LNCSE, pages 221–228. Springer-Verlag, 2021.

[134] V. Dolean, P. Jolivet, and F. Nataf. *An Introduction to Domain Decomposition Methods*. Society for Industrial and Applied Mathematics, 2015.

[135] W. Dörfler and M. Ainsworth. Reliable a posteriori error control for nonconforming finite element approximation of Stokes flow. *Mathematics of Computation*, 74(252):1599–1619, 2005.

[136] J. Droniou, R. Eymard, T. Gallouët, C. Guichard, and R. Herbin. *The Gradient Discretization Method*. Springer, 2018.

[137] H. Duan, Z. Du, W. Liu, and S. Zhang. New mixed element for Maxwell equations. *SIAM Journal on Numerical Analysis*, 57:320–354, 2019.

[138] H. Duan, F. Jia, P. Lin, and R. C. E. Tan. The local L^2 projected C^0 finite element method for Maxwell problem. *SIAM Journal on Numerical Analysis*, 47:1274–1303, 2009.

[139] H. Duan, S. Li, R. C. E. Tan, and W. Zheng. A delta-regularization finite element method for a double curl problem with divergence-free constraint. *SIAM Journal on Numerical Analysis*, 50:3208–3230, 2012.

[140] F. Dubois and O. Lafitte. An analytic and symbolic analysis of a coupled thermo-neutronic problem. In *2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 61–65, 2021.

[141] F. Dubois, M. Salaün, and S. Salmon. First vorticity–velocity–pressure numerical scheme for the Stokes problem. *Computer Methods in Applied Mechanics and Engineering*, 192(44):4877–4907, 2003.

[142] F. Dubois, M. Salaün, and S. Salmon. Vorticity–velocity-pressure and stream function-vorticity formulations for the Stokes problem. *Journal de Mathématiques Pures et Appliquées*, 82(11):1395–1451, 2003.

[143] J. J. Duderstadt and L. J. Hamilton. *Nuclear reactor analysis*. John Wiley & Sons, Inc., 1976.

[144] S. Eddargani. *Approximation by splines functions on Powell-Sabin triangulations*. PhD thesis, Granada University and Hassan 1er University, 2021.

[145] P. Emonot. *Méthodes de volumes éléments finis : applications aux équations de Navier-Stokes et résultats de convergence*. PhD thesis, Université Lyon I, 1992.

[146] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*. Springer-Verlag, 2004.

[147] A. Ern and J.-L. Guermond. *Finite elements I*, volume 72 of *Texts in Applied Mathematics*. Springer, 2021.

[148] A. Ern and J.-L. Guermond. *Finite elements II*, volume 73 of *Texts in Applied Mathematics*. Springer, 2021.

[149] M. Fabien, J. Guzmán, M. Neilan, and A. Zytoon. Low-order divergence-free approximations for the Stokes problem on Worsey-Farin and Powell-Sabin splits. *Computer Methods in Applied Mechanics and Engineering*, 390:114444, 2022.

[150] M. Feistauer and A. Ženišek. Finite element solution of nonlinear elliptic problems. *Numer. Math.*, 50:451–475, 1987.

[151] P. Fernandes and G. Gilardi. Magnetostatic and electrostatic problems in inhomogeneous anisotropic media with irregular boundary and mixed boundary conditions. *Math. Models Meth. App. Sci.*, 7(7):957–991, 1997.

[152] M. Fortin and M. Soulie. A non-conforming piecewise quadratic finite element on triangles. *International Journal for Numerical Methods in Engineering*, 19(4):505–520, 1983.

[153] T. Fortin. *Une méthode éléments finis à décomposition L^2 d'ordre élevé motivée par la simulation d'écoulement diphasique bas Mach*. PhD thesis, Université Paris VI, 2006.

[154] D. Gallistl. Rayleigh-Ritz approximation of the inf-sup constant for the divergence. *Mathematics of Computation*, 88(315):73–89, 2019.

[155] M. J. Gander and C. Japhet. An algorithm for non-matching grid projections with linear complexity. In *Domain Decomposition Methods in Science and Engineering XVIII, Lecture Notes in Computational Science and Engineering*, pages 185–192. Springer-Verlag, 2009.

[156] M. J. Gander and C. Japhet. Algorithm 932: PANG: Software for Non-Matching Grid Projections in 2d and 3d with Linear Complexity. *ACM Transactions on Mathematical Software (TOMS)*, 40(1):6:1–6:25, 2013.

[157] E. Garcia. *Résolution des équations de Maxwell instationnaires avec charges dans des domaines non convexes*. PhD thesis, Université Paris VI, 2002.

[158] G. Gatica. *A Simple Introduction to the Mixed Finite Element Method: Theory and Applications*. SpringerBriefs in Mathematics. Springer, 2014.

[159] E. Gelbard. Application of Spherical Harmonics Method to Reactor Problems. Report WAPD-BT-20, Westinghouse, 1960.

[160] M. Gervais, F. Madiot, M.-H. Do, and P. Ciarlet Jr. Adaptive solution of the domain decomposition+L2-jumps method applied to the neutron diffusion equation on structured meshes. *EPJ Web Conf.*, 302:02011, 2024.

[161] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations*. Springer-Verlag, 1986.

[162] V. Girault, B. Rivière, and M. F. Wheeler. A splitting method using discontinuous Galerkin for the transient incompressible Navier-Stokes equations. *ESAIM: M2AN*, 39(6):1115–1147, 2005.

[163] L. Giret. *Non-Conforming Domain Decomposition for the Multigroup Neutron SPN Equation*. PhD thesis, Université Paris-Saclay, 2018.

[164] H. Golfier, R. Lenain, J.-J. Lautard, P. Fougeras, P. Magat, E. Martinolli, and Y. Duteillet. APOLLO3: a common project of CEA, AREVA and EDF for the development of a new deterministic multi-purpose code for core physics analysis. In *International Conference on Advances in Mathematics, Computational Methods, and Reactor Physics (M&C 2009)*. American Nuclear Society (ANS), 2009.

[165] S. Gounand. https://cea.hal.science/cea-02489502/document, 2015.

[166] P. Grisvard. *Elliptic problems in nonsmooth domains*. Pitman, 1985.

[167] P. Grisvard. *Singularities in boundary value problems*, volume 22 of *RMA*. Masson, 1992.

[168] P. W. Gross and P. R. Kotiuga. *Electromagnetic theory and computation: a topological approach*, volume 48 of *MSRI Publications Series*. Cambridge University Press, Cambridge, 2004.

[169] J.-L. Guermond and P. D. Minev. High-Order Time Stepping for the Incompressible Navier–Stokes Equations. *SIAM Journal on Scientific Computing*, 37(6):A2656–A2681, 2015.

[170] J.-L. Guermond and P. D. Minev. High-order time stepping for the Navier–Stokes equations with minimal computational complexity. *Journal of Computational and Applied Mathematics*, 310:92–103, 2017.

[171] M. Halla. Galerkin approximation of holomorphic eigenvalue problems: weak T-coercivity and T-compatibility. *Numer. Math.*, 148:387–407, 2021.

[172] F. H. Harlow and J. E. Welch. Numerical calculation of time-dependent viscous incompressible flow of fluid with a free surface. *Physics of Fluids*, 8(2):2182–2189, 1965.

[173] C. Hazard and S. Lohrengel. A singular field method for Maxwell's equations: numerical aspects for 2D magnetostatics. *SIAM Journal on Applied Mathematics*, 40:1021–1040, 2002.

[174] F. Hecht. Construction d'une base de fonctions P_1 non conforme à divergence nulle dans $\mathbb{R}^3$. *RAIRO, Sér. Anal. Numé.*, 15(2):119–150, 1981.

[175] S. Heib. *Nouvelles discrétisations non structurées pour des écoulements de fluides à incompressibilité renforcée*. PhD thesis, Université Paris VI, 2003.

[176] R. Hiptmair. Finite elements in computational electromagnetics. *Acta Numerica*, 50:237–339, 2002.

[177] T. Hohage and L. Nannen. Convergence of infinite element methods for scalar waveguide problems. *BIT Numer. Math.*, 55:215–254, 2015.

[178] H. Hutridurga, O. Mula, and F. Salvarani. Homogenization in the energy variable for a neutron transport model. *Asymptotic Analysis*, 117(1–2):1–25, 2020.

[179] D. S. Jerison and C. E. Kenig. The Neumann problem on Lipschitz domains. *Bull. Amer. Math. Soc.*, 4:203–207, 1981.

[180] V. John, A. Linke, C. Merdon, M. Neilan, and L.G. Rebholz. On the Divergence Constraint in Mixed Finite Element Methods for Incompressible Flows. *SIAM REVIEW*, 59(3):492–544, 2017.

[181] F. D. Kpadonou. *Shape and anisotropy optimization by an isogeometric-polar method*. PhD thesis, Université Paris-Saclay, 2017.

[182] J. Krebs, M.-C. Laigle, R. Lenain, G. Mathonnière, and A. Nicolas. Calculational methods for PWR. In *Specialist Meeting on Advanced Calculational Methods for Power Reactors, Cadarache, France*. International Atomic Energy Agency, 1990.

[183] M. G. Krein and M. A. Rutman. *Linear operators leaving invariant a cone in a Banach space*. Amer. Math. Soc. Translation, 1950.

[184] L. Cattabriga. Su un problema al contorno relativo al sistema di equazioni di Stokes. *Rendiconti del Seminario Matematico della Università di Padova*, 31:308–340, 1961.

[185] J.-J. Lautard, S. Loubière, and C. Fedon-Magnaud. Cronos: a modular computational system for neutronic core calculations. Technical report, CEA/DANS/DEN/DM2S/SERMA, 1992.

[186] C. Le Potier. A finite volume method for the approximation of highly anisotropic diffusion operators on unstructured meshes. In *Finite Volumes for Complex Applications IV, Marrakesh, Marocco*, 2005.

[187] C. Le Potier. Schéma volumes finis pour des opérateurs de diffusion fortement anisotropes sur des maillages non structurés. *Comptes Rendus Mathematique*, 341(12):787–792, 2005.

[188] C. Le Potier. *Construction et développement de nouveaux schémas numériques pour des problèmes elliptiues ou paraboliques.* Habilitation à diriger des recherches, Université Paris-Est, 2017.

[189] B. V. Leer. Towards the ultimate conservative difference scheme. A second-order sequel to Godunov's method. *Journal of Computational Physics*, 32:101–136, 1979.

[190] A. Lefebvre-Lepot and F. Nabet. Numerical simulation of rigid particles in Stokes flow: lubrication correction for general shape of particles. *Math. Model. Nat. Phenom.*, 16:45, 2021.

[191] F. Lefèvre, S. Lohrengel, and S. Nicaise. An extended finite element method for 2d edge elements. *Int. J. Numer. Anal. Model.*, 8(4):641–666, 2011.

[192] P. Lesaint and P.-A. Raviart. On a finite element method for solving the neutron transport equation. *Publications mathématiques et informatique de Rennes*, S4:1–40, 1974.

[193] Y. Li and L. T. Zikatanov. New stabilized $P_1 \times P_0$ finite element methods for nearly inviscid and incompressible flows. *Computer Methods in Applied Mechanics and Engineering*, 393:114815, 2022.

[194] A. Linke. On the Role of the Helmholtz-Decomposition in Mixed Methods for Incompressible Flows and a New Variational Crime. *Comput. Methods Appl. Mech. Engrg.*, 268:782–800, 2014.

[195] P.-L. Lions. On the Schwarz alternating method I. In R. Glowinski et al., editors, *First International Symposium on Domain Decomposition Methods for Partial Differential equations*, pages 1–42. SIAM, 1988.

[196] P.-L. Lions. On the schwarz alternating method III: a variant for nonoverlapping subdomains. In T. Chan et al., editors, *Third International Symposium Domain Decomposition Methods for Partial Differential equations*, pages 202–223. SIAM, 1990.

[197] K. Lipnikov, M. Shashkov, and I. Yotov. Local flux mimetic finite difference methods. *Numer. Math.*, 112:115–152, 2009.

[198] J.-J. Moller, J.-Y. Lautard. Minaret, a deterministic neutron transport solver for nuclear core calculations. In *M&C 2011*. International Atomic Energy Agency, 2011.

[199] P. Monk. *Finite element methods for Maxwell's equations.* Oxford University Press, Oxford, Great Britain, 2003.

[200] O. Mula. *Some contributions towards the parallel simulation of time dependent neutron transport and the integration of observed data in real time.* PhD thesis, Université Paris VI, 2014.

[201] F. Nataf and F. Nier. Convergence rate of some domain decomposition methods for overlapping and nonoverlapping subdomains. *Numer. Math.*, 75:357–377, 1997.

[202] S. Nicaise and A.-M. Sändig. General interace problems-i-ii. *Math. Methods Appl. Sci.*, 17(6):395–450, 1994.

[203] S. Nicaise and J. Venel. A posteriori error estimates for a finite element approximation of transmission problems with sign changing coefficients. *J. Comput. Appl. Math.*, 235:4272–4282, 2011.

[204] J.-C. Nédélec. Mixed finite elements in $\mathbb{R}^3$. *Numer. Math.*, 35:315–341, 1980.

[205] M. A. Olshanskii and E. E. Tyrtyshnikov. *Iterative Methods for Linear Systems.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[206] J. E. Osborn. Spectral approximation for compact operators. *Math. Comp.*, 29:712–725, 1975.

[207] R. Otin. Regularized Maxwell equations and nodal finite elements for electromagnetic field computations. *Electromagnetics*, 30:190–204, 2010.

[208] R. Otin. *Regularized Maxwell equations and nodal finite elements for electromagnetic field computations in frequence domain.* PhD thesis, Polytechnic University of Catalonia, 2011.

[209] R. Otin. ERMES: A nodal-based finite element code for electromagnetic simulations in frequency domain. *Computer Physics Communications*, 184:2588–2595, 2013.

[210] A. Peitavy. *Schémas de discrétisation en pression et éléments finis de Crouzeix-Raviart pour les écoulements de fluides incompressibles.* PhD thesis, Université Gustave Eiffel, 2024.

[211] P. Percell. On cubic and quartic Clough-Tocher finite elements. *SIAM Journal on Numerical Analysis*, 13:100–103, 1976.

[212] A. Quarteroni and A. Valli. *Domain decomposition methods for partial differential equations*. Oxford Science Publications, 1999.

[213] P.-A. Raviart and J.-M. Thomas. A mixed finite element method for second order elliptic problems. In *Mathematical aspects of finite element methods*, volume 606 of *Lecture Notes in Mathematics*, pages 292–315. Springer, 1977.

[214] P.-A. Raviart and J.-M. Thomas. *Introduction à l'Analyse Numérique des Équations aux Dérivées Partielles*. Masson, 1983.

[215] W. H. Reed and T. R. Hill. Triangular mesh methods for the neutron transport equation. Technical Report LA-UR-73-479, Los Alamos Scientific Lab., N.Mex. (USA), 1973.

[216] S. Reitzinger and J. Schöberl. An algebraic multigrid method for finite element discretizations with edge elements. *Numer. Linear Algebra Appl.*, 9:223–238, 2002.

[217] J. E. Roberts and J.-M. Thomas. Mixed and hybrid methods. In *Finite Element Methods (Part 1)*, volume 2 of *Handbook of Numerical Analysis*, pages 523–629. Elsevier, 1991.

[218] E. Rudoy. Domain decomposition method for crack problems with nonpenetration condition. *ESAIM: M2AN*, 50(4):995–1009, 2016.

[219] S. Gong and M. J. Gander and I G. Graham and D. Lafontaine and E. A. Spence. Convergence of parallel overlapping domain decomposition methods for the Helmholtz equation. *Numer. Math.*, 152(2):259–306, 2022.

[220] S. Lohrengel and S. Nicaise. Analysis of eddy current formulations in two-dimensionel domains with cracks. *ESAIM: M2AN*, 49:141–170, 2015.

[221] S. Zhang. A new family of stable mixed finite elements for the 3D Stokes equations. *Mathematics of Computation*, 74:543–554, 2005.

[222] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on scientific and statistical computing*, 7(3):856–869, 1986.

[223] M. Salaün and S. Salmon. Numerical stabilization of the Stokes problem in vorticity–velocity–pressure formulation. *Computer Methods in Applied Mechanics and Engineering*, 196(9):1767–1786, 2007.

[224] M. Salaün and S. Salmon. Low-order finite element method for the well-posed bidimensional Stokes problem. *IMA Journal of Numerical Analysis*, 35(1):427–453, 2014.

[225] A. Sargeni, K. W. Burn, and G. B. Bruna. Coupling effects in large reactor cores: the impact of heavy and conventional reflectors on power distribution perturbations. In *PHYSOR 2014, Kyoto, Japan, Sept 28-Oct 3, 2014*, 2014.

[226] S. Sauter. The inf-sup constant for *hp*-Crouzeix-Raviart triangular elements, 2022.

[227] S. Sauter. The inf-sup constant for *hp*-Crouzeix-Raviart triangular elements. *Computers and Mathematics with Applications*, 149:49–70, 2023.

[228] S. Sauter and C. Torres. On the Inf-Sup Stabillity of Crouzeix-Raviart Stokes Elements in 3D. *Mathematics of Computation*, 92:1033–1059, 2023.

[229] H. A. Schwarz. Uber einige Abbildungsaufgaben. *Ges. Math Abh.*, 11:65–83, 1869.

[230] L. R. Scott and M. Vogelius. Norm estimates for a maximal right inverse of the divergence operator in spaces of piecewise polynomials. *RAIRO, Sér. Anal. Numér.*, 19(1):111–143, 1985.

[231] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54:483–493, 1990.

[232] J. Shen. On error estimates of the penalty method for unsteady navier-stokes equations. *SIAM Journal on Numerical Analysis*, 32(2):386–403, 1995.

[233] T. Sorokina and A. J. Worsey. A multivariate Powell-Sabin interpolant. *Adv Comput Math*, 29:71–89, 2008.

[234] R. Stenberg. Error analysis of some finite element methods for the Stokes problem. *Math. Comp.*, 54:495–508, 1990.

[235] T. Takeda and H. Ikeda. 3-D neutron transport benchmarks. *J. Nucl. Sci. Technol.*, 28(7):656–669, 1991.

[236] Z. Tang, Y. Le Menach, E. Creusé, S. Nicaise, F. Piriou, and N. Nemitz. Residual and equilibrated error estimators for magnetostatic problems solved by Finite Element Method. *IEEE Trans. Magn.*, 49:5715–5723, 2013.

[237] L. Tartar. *An introduction to Sobolev spaces and interpolation spaces*, volume 3 of *Lecture Notes of the Unione Matematica Italiana*. Springer, 2007.

[238] Y. C. Taumhas, D. Labeurthre, F. Madiot, O. Mula, and T. Taddei. Impact of physical model error on state estimation for neutronics applications. *ESAIM: ProcS*, 73:158–172, 2023. CEMRACS 2021 - Data Assimilation and Reduced Modeling for High Dimensional Problems.

[239] C. Taylor and T. Hood. Numerical solution of the Navier-Stokes equations using the finite element technique. *Computers & Fluids*, 1:73–100, 1973.

[240] R. Temam. Une méthode d'approximation de la solution des équations de Navier-Stokes. *Bulletin de la Société Mathématique de France*, 96:115–152, 1968.

[241] G. Unger. Convergence analysis of a Galerkin boundary element method for electromagnetic resonance problems. *SN Partial Differential Equations and Applications*, 39, 2021.

[242] M. Vohralík. On the discrete Poincaré-Friedrichs inequlities for nonconforming approximations of the Soboled space H^1. *Numer. Funct. Anal. Optim.*, 26:925–952, 2005.

[243] T. Warburton and J. S. Hesthaven. On the constants in hp-finite element trace inverse inequalities. *Comput. Methods Appl. Mech. Engrg.*, 192(25):2765–2773, 2003.

[244] C. Weber. A local compactness theorem for Maxwell's equations. *Mathematical Methods in the Applied Sciences*, 2(1):12–25, 1980.

[245] M. F. Wheeler and I. Yotov. A posteriori error estimates for the mortar mixed finite element method. *SIAM Journal on Numerical Analysis*, 43(3):1021–1042, 2005.

[246] B. I. Wohlmuth. *Discretization methods and iterative solvers based on domain decomposition*, volume 17 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, 2001.

[247] I. Yotov. *Mixed Finite Element Methods for Flow in Porous Media*. PhD thesis, Rice University, Houston, USA, 1996.

[248] A.-A. Zepernick. A survey on interpolation and projection operators. Master's thesis, Free university of Berlin, 2023.

[249] T. Zhang and L. Tang. A stabilized finite volume method for Stokes equations using the lowest order $P_1 - P_0$ element pair. *Adv Comput Math*, 41:781–798, 2015.